Progress Toward Transforming the Integrated Risk Information System (IRIS) Program: A 2018 Evaluation (2018)

DETAILS

129 pages | 8.5 x 11 | PAPERBACK
ISBN 978-0-309-47491-7 | DOI 10.17226/25086

CONTRIBUTORS

Committee to Review Advances Made to the IRIS Process; Board on Environmental Studies and Toxicology; Division on Earth and Life Studies; National Academies of Sciences, Engineering, and Medicine

SUGGESTED CITATION

National Academies of Sciences, Engineering, and Medicine 2018. *Progress Toward Transforming the Integrated Risk Information System (IRIS) Program: A 2018 Evaluation*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25086.

Progress Toward Transforming the Integrated Risk Information System (IRIS) Program: A 2018 Evaluation

Committee to Review Advances Made to the IRIS Process

Board on Environmental Studies and Toxicology

Division on Earth and Life Studies

A Consensus Study Report of
The National Academies of
SCIENCES • ENGINEERING • MEDICINE

THE NATIONAL ACADEMIES PRESS
Washington, DC
www.nap.edu

THE NATIONAL ACADEMIES PRESS **500 Fifth Street, NW** **Washington, DC 20001**

This project was supported by Contract EP-C-14-005, TO#008 between the National Academies of Sciences, Engineering, and Medicine and the US Environmental Protection Agency. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the view of any organization or agency that provided support for this project.

Digital Object Identifier: https://doi.org/10.17226/25086

Printed in the United States of America

Suggested citation: National Academies of Sciences, Engineering, and Medicine. 2018. *Progress Toward Transforming the Integrated Risk Information System (IRIS) Program: A 2018 Evaluation*. Washington, DC: The National Academies Press. doi: https://doi.org/10.17226/25086.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, nongovernmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia McNutt is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. C. D. Mote, Jr., is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the **National Academies of Sciences, Engineering, and Medicine** to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The National Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at **www.nationalacademies.org**.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Consensus Study Reports published by the National Academies of Sciences, Engineering, and Medicine document the evidence-based consensus on the study's statement of task by an authoring committee of experts. Reports typically include findings, conclusions, and recommendations based on information gathered by the committee and the committee's deliberations. Each report has been subjected to a rigorous and independent peer-review process and it represents the position of the National Academies on the statement of task.

Proceedings published by the National Academies of Sciences, Engineering, and Medicine chronicle the presentations and discussions at a workshop, symposium, or other event convened by the National Academies. The statements and opinions contained in proceedings are those of the participants and are not endorsed by other participants, the planning committee, or the National Academies.

For information about other products and activities of the National Academies, please visit www.nationalacademies.org/about/whatwedo.

COMMITTEE TO REVIEW ADVANCES MADE TO THE IRIS PROCESS

Members

JONATHAN M. SAMET (*Chair*), Colorado School of Public Health, Denver, CO
RICHARD A. CORLEY (retired), Pacific Northwest National Laboratories, Richland, WA
GEORGE DASTON, Proctor & Gamble Company, Cincinnati, OH
DAVID C. DORMAN, North Carolina State University, Raleigh, NC
RUSS B. HAUSER, Harvard T.H. Chan School of Public Health, Boston, MA
KAREN A. ROBINSON, Johns Hopkins University, Baltimore, MD
RICHARD P. SCHEINES, Carnegie Mellon University, Pittsburgh, PA
LAUREN ZEISE, CalEPA Office of Environmental Health Hazard Assessment, Berkeley, CA
YILIANG ZHU, University of New Mexico, Albuquerque, NM

Staff

ELLEN K. MANTUS, Project Director
RADIAH ROSE-CRAWFORD, Manager, Editorial Projects
SUZANNE THILENIUS, Administrative Coordinator
MIRSADA KARALIC-LONCAREVIC, Manager, Technical Information Center
JESSICA WOLFMAN, Senior Program Assistant

Sponsor

US ENVIRONMENTAL PROTECTION AGENCY

Acknowledgments

This Consensus Study Report was reviewed in draft form by persons chosen for their diverse perspectives and technical expertise. The purpose of this independent review is to provide candid and critical comments that will assist the National Academies of Sciences, Engineering, and Medicine in making each published report as sound as possible and to ensure that it meets institutional standards of quality, objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process.

We thank the following for their review of this report:

David Eaton, University of Washington
David Savitz, Brown University
Joyce Tsuji, Exponent
Harvey Clewell, Ramboll
Lisa Bero, University of Sidney
Nancy Reid, University of Toronto

Although the reviewers listed above provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations of this report, nor did they see the final draft before its release. The review of the report was overseen by Mark Cullen, Stanford University, who was responsible for making certain that an independent examination of the report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content rests entirely with the authoring committee and the National Academies.

The committee gratefully acknowledges the staff of the US Environmental Protection Agency, especially Tina Bahadori and Kristina Thayer, for their presentations to the committee during open sessions. The committee is also grateful for the assistance of Norman Grossblatt who served as the report editor.

Contents

Summary

Over the past several years, the US Environmental Protection Agency (EPA) has been transforming the procedures of its Integrated Risk Information System (IRIS), a program that produces hazard and dose–response assessments of environmental chemicals and derives toxicity values that can be used to estimate risks posed by exposures to them. The transformation was initiated after suggestions for program reforms were provided in a 2011 report from the National Academies of Sciences, Engineering, and Medicine that reviewed a draft IRIS assessment of formaldehyde. In 2014, the National Academies released a report that reviewed the IRIS program and evaluated the changes implemented in it since the 2011 report. Although it provided many recommendations, the 2014 report concluded that "substantial improvements in the IRIS process have been made, and it is clear that EPA has embraced and is acting on the [National Academies] recommendations."

Since 2014, new leadership of EPA's National Center for Environmental Assessment (NCEA) and IRIS program has instituted even more substantive changes in the IRIS program in response to the recommendations in the 2014 report. Given the new direction of the IRIS program, EPA asked the National Academies to review the agency's progress toward addressing the past recommendations. Accordingly, the National Academies convened the Committee to Review Advances Made to the IRIS Process. The present committee heard presentations, reviewed posters, and received demonstrations of toolkits and databases from EPA over the course of a 1.5-day workshop, and it reviewed recent IRIS work products. This brief report provides the committee's general findings regarding EPA's progress (Chapter 2) and specific findings regarding changes made in response to individual recommendations from the 2014 report (Appendix E).

Overall, the committee was impressed with the changes being instituted in the IRIS program since the 2014 report. The committee views the transformation of the IRIS program as a work in progress, recognizes that this review assesses one moment in time in a still-evolving program, and acknowledges that the IRIS program will (and should) continue to evolve as it adapts and applies new scientific approaches and knowledge. The change in NCEA and IRIS leadership has led to substantive reforms, and there is strong evidence that systematic review methods are being developed and implemented and that there is a commitment to use systematic-review methods to conduct IRIS assessments. Although the committee offers some refinements and identifies a few possibilities for further development in Chapter 2, its overall conclusion is that EPA has been responsive and has made substantial progress in implementing National Academies recommendations.

1

Introduction

For many years, the National Academies of Sciences, Engineering, and Medicine has been asked to review assessments produced by the Integrated Risk Information System (IRIS) of the US Environmental Protection Agency (EPA). The reviews have consistently provided recommendations for revisions of specific assessments, but the National Academies committee that was tasked with reviewing the draft IRIS assessment of formaldehyde also suggested changes to improve the IRIS program itself, if EPA chose to do so. Since release of that committee's report (NRC 2011), the IRIS program has been undergoing substantive changes. In 2014, another National Academies committee reviewed the changes in the IRIS program and provided an overall favorable assessment, noting that it was reviewing a work in progress (NRC 2014). In light of a change in leadership and continued revisions of the IRIS program, EPA asked the National Academies to review changes since 2014 and to determine whether they have been responsive to the recommendations in past National Academies reports. In response to EPA's request, the National Academies convened the Committee to Review Advances Made to the IRIS Process, which prepared this brief report.

THE INTEGRATED RISK INFORMATION SYSTEM AND PREVIOUS NATIONAL ACADEMIES REPORTS

Given problems in several IRIS assessments noted by previous National Academies committees (see, for example, NRC 2006, 2010, 2011) and specific issues encountered in the formaldehyde assessment, the committee that evaluated the formaldehyde assessment provided a roadmap for reframing the development of IRIS assessments (Chapter 7, NRC 2011). The roadmap did not provide detailed guidance but rather suggestions for creating a more systematic and transparent IRIS process, if EPA chose to go forward with reforming the process. Congress directed EPA to respond to and incorporate the recommendations and suggestions provided in Chapter 7 of the 2011 National Academies report (House Report 112-151; Public Law 112-74). EPA indicated that the agency was committed to responding to National Academies recommendations and improving the IRIS program and began to make substantive changes. In a 2012 report to Congress, EPA highlighted its intended changes, such as a new document structure with a preamble that describes general methods for evidence identification, evidence evaluation, and derivation of toxicity values; new systematic approaches for data analysis; and expanded efforts for stakeholder engagement (EPA 2012; NRC 2014). EPA also noted that it had formed the Chemical Assessment Advisory Committee under the auspices of its Scientific Advisory Board to advise the agency on specific assessments and broader program issues. To ensure that EPA was responding adequately to National Academies recommendations, Congress asked the National Academies to review the changes that EPA was implementing.

In 2014, the National Academies released the report *Review of EPA's Integrated Risk Information System (IRIS) Process* (NRC 2014), which evaluated the changes that were being implemented in the IRIS program and assessed whether they were responsive to the recommendations and suggestions made in Chapter 7 of the 2011 report. The 2014 report concluded that "substantial improvements in the IRIS process have been made, and it is clear that EPA has embraced and is acting on the recommendations in the…formaldehyde report." It urged EPA to adopt systematic review practices, framed the IRIS process

in the context of systematic review (see Figure 1-1), and provided specific recommendations on each step of the process (NRC 2014). Since release of the 2014 report, substantive efforts have been made to incorporate systematic review into the IRIS process, and EPA has now asked the National Academies to assess its progress.

THE COMMITTEE, ITS TASK, AND ITS APPROACH

The committee that was convened to address EPA's request included expertise in toxicology, epidemiology, risk assessment, statistics, modeling, evidence integration, and systematic review; see Appendix A for biographic information on the committee. The verbatim statement of the committee's task is provided in Box 1-1. As noted, the committee was asked to assess the changes that have been (or that are planned to be) implemented by EPA in response to National Academies recommendations. It is important to note that the committee was not asked to evaluate the overall value of the IRIS program, to recommend where IRIS should be located within the agency, or to review any specific chemical assessment. The committee was also not tasked with conducting a comprehensive review of the IRIS program; rather, it was asked to evaluate whether the current trajectory of the program agrees with past recommendations of the National Academies.

To address its task, the committee held a 1.5-day workshop during which EPA presented its progress to the committee. Multiple opportunities for stakeholder input were provided. Appendix B provides the workshop agenda. The committee reviewed EPA presentations (Appendix C), posters (Appendix D), recently released materials (EPA 2017, 2018a,b; Orme-Zavaleta 2018), and all materials submitted by stakeholders. To fulfill its task of evaluating EPA's progress in implementing past National Academies recommendations, the committee decided to focus its attention primarily on recommendations made in the report *Review of EPA's Integrated Risk Information System (IRIS) Process* (NRC 2014). Although the report *Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde* (NRC 2011) provided general suggestions for reforming the IRIS program, it primarily made recommendations specifically for revising the draft assessment of formaldehyde. It is important to note that the 2011 committee was not tasked with an extensive review of the IRIS program. The 2014 report considered the general suggestions provided in the 2011 report, reviewed the IRIS program specifically, and made numerous recommendations directed at the program. Therefore, the present committee considered the 2014 report as expanding on the suggestions provided in the 2011 report and thus evaluated EPA's progress in addressing each recommendation in the 2014 report.

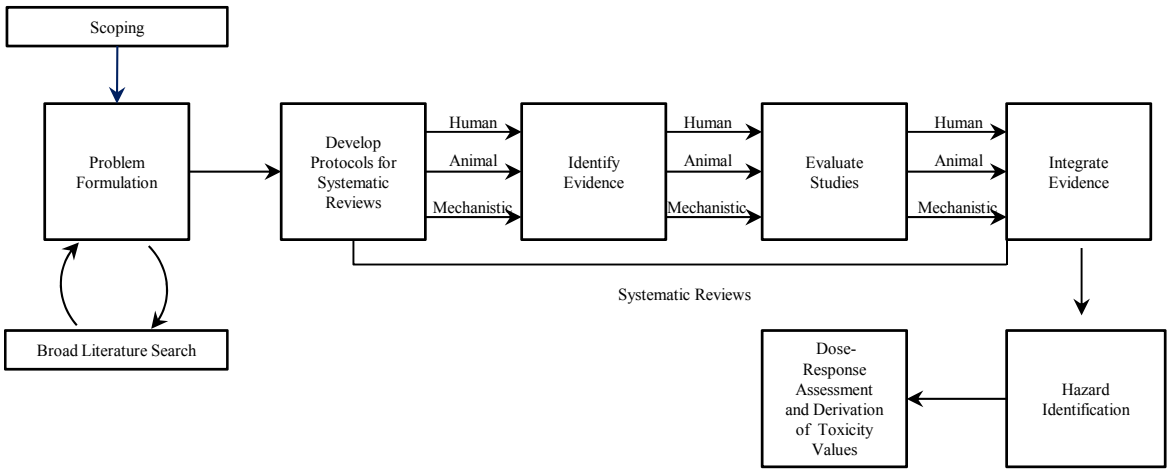


FIGURE 1-1 The IRIS process in the context of systematic review. Source: NRC (2014).

BOX 1-1 Statement of Task

An ad hoc committee under the auspices of the National Academies of Sciences, Engineering, and Medicine will assess changes that have been implemented or plan to be implemented by the U.S. Environmental Protection Agency (EPA) for its Integrated Risk Information System (IRIS) in response to recommendations made in previous NRC reports, such as *Review of EPA's Integrated Risk Information System (IRIS) Process* and *Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde*. The committee will base its assessment on EPA presentations and interactive sessions during a 1.5 day workshop at which multiple opportunities will be provided for stakeholder input.

ORGANIZATION OF THE REPORT

The present report is organized into two chapters and five appendixes. Chapter 2 presents the committee's overall findings regarding advances made in the IRIS process. Appendix A provides biographic information on the committee. Appendixes B, C, and D provide the workshop agenda, EPA presentations made during the workshop, and EPA poster presentations, respectively. Appendix E details the committee's findings concerning individual recommendations in the report *Review of EPA's Integrated Risk Information System (IRIS) Process* (NRC 2014).

REFERENCES

EPA (US Environmental Protection Agency). 2012. EPA's Integrated Risk Information Program, Progress Report and Report to Congress. Office of Research and Development, U.S. Environmental Protection Agency [online]. Available: https://www.epa.gov/sites/production/files/2015-06/documents/iris_report_to_congress_2012.pdf [accessed February 27, 2018].

EPA. 2017. IRIS Assessment Plan for Chloroform [CASRN 67-66-3]. EPA/635/R-17/330. Integrated Risk Information System, National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC.

EPA. 2018a. Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation) [CASRN 67-66-3]. EPA/635/R-17/486. Integrated Risk Information System, National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC.

EPA. 2018b. IRIS Assessment Plan for Uranium (Oral Reference Dose) (Scoping and Problem Formulation Materials) [CASRN 7440-61-1]. EPA/635/R-17/787. Integrated Risk Information System, National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC.

NRC (National Research Council). 2006. Health Risks from Dioxin and Related Compounds: Evaluation of the EPA Reassessment. Washington, DC: The National Academies Press.

NRC. 2010. Review of the Environmental Protection Agency's Draft IRIS Assessment of Tetrachloroethylene. Washington, DC: The National Academies Press.

NRC. 2011. Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde. Washington, DC: The National Academies Press.

NRC. 2014. Review of EPA's Integrated Risk Information System (IRIS) Process. Washington, DC: The National Academies Press.

Orme-Zavaleta, J. 2018. Response to the Request for Correction (RFC). Letter to Robert Holden, Liskow & Lewis, New Orleans, LA, from Jennifer Orme-Zavaleta, Principal Deputy Assistant Administrator for Science, Office of Research and Development, Washington, DC, January 25, 2018; Attachment 1. EPA Response to the Denka Performance Elastomers (DPE) Request for Correction of the Toxicological Review of Chloroprene (CAS No. 126-99-8) In Support of Summary Information on the Integrated Risk Information System (IRIS); Attachment 2. Systematic Review of Chloroprene [CASRN 126-99-80] Studies Published Since 2010 IRIS Assessment to Support Consideration of the Denka Request for Correction (RFC). January 2018 [online]. Available: https://www.epa.gov/sites/production/files/2018-01/documents/epa_repsonse_to_mr._holdren_jan_25_2018_complete.pdf [accessed February 9, 2018].

2

Responses to National Academies Recommendations

Over the course of a 1.5-day workshop, the US Environmental Protection Agency (EPA) made presentations to the committee on changes that are transforming the Integrated Risk Information System (IRIS) process. The committee used that information and recently released IRIS documents to judge the extent to which EPA has adequately addressed recommendations made in previous National Academies reports, primarily *Review of EPA's Integrated Risk Information System (IRIS) Process* (NRC 2014).[1] The committee's overall comments are provided below; findings regarding individual recommendations are in Appendix E.

GENERAL PROCESS ISSUES

The 2014 report (Chapter 2 in NRC 2014) offered recommendations related to the IRIS process and evaluated EPA's progress in implementing the suggestions made in *Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde* (NRC 2011).[2] Above all, the 2014 report commented on the need to continue to sustain the evolution of the program's procedures and to consider how EPA will do so in the context of continually advancing scientific methods. In the 4 years since the 2014 recommendations, the IRIS program clearly has maintained a trajectory of change that has accelerated under the new leadership of the EPA National Center for Environmental Assessment (NCEA) and the IRIS program. The committee was impressed by the scope of changes that have been or are being implemented and by the engagement of scientists throughout NCEA, EPA more broadly, other federal agencies, and academe to effect change. Such engagement is appropriate inasmuch as funding for the use of external contractors has diminished, and there is expertise in relevant fields throughout the agency. Supervisory and communication strategies are in place, and formal quality management has been implemented. The committee notes that EPA will need to ensure that quality management extends to activities that are conducted by people who are outside the IRIS program.

Changes in some of the critical elements of the overall IRIS process are still in progress. The 2014 committee was given an incomplete draft of the handbook; the handbook is intended to provide guidance on the IRIS process. The 2014 committee recommended completion of that handbook; the present committee was not given a draft of the handbook. EPA indicated that the handbook is still in development and is "being updated to reflect Agency input, evolving IRIS practices as systematic-review approaches are tested through implementation, and public comment received on chemical-specific protocols" (Slide 22, Appendix C). Public release is anticipated in 2018. The handbook is expected to provide critical guidance for the development of IRIS assessments, and the present committee urges that high priority be given to its completion, peer review, and release. Reference to it will facilitate transparency on the approach for specific IRIS assessments. In the absence of a final version of the handbook, EPA is describing its approach for the reviews in its protocol documents, and this practice provides transparency into the assessment process while the handbook is being completed. The committee notes, however, that the handbook should not become a final, fixed set of guidelines but rather should be a document that evolves over time.

[1]Referred to hereafter as the 2014 report. The committee that produced that report is referred to as the 2014 committee.

[2]Referred to hereafter as the 2011 report.

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

The 2014 committee also commented on the need to incorporate input from various stakeholders—including industry, academe, and nongovernment organizations—at appropriate points in the process; this recommendation has been heeded by past and current program leaders. Three points in the process, including development of assessment plans and systematic-review protocols, have been identified at which public comments will be sought (slide 24, Appendix C). Although the present committee was not shown the approach for acknowledging public comments and incorporating them into the process, the handbook should describe how this will be done. The committee was impressed by other NCEA program activities that engage stakeholders, including dissemination of tools that it has developed, such as the benchmark-dose (BMD) modeling software, and provision of training.

EPA also commented that it was moving away from a one-size-fits-all approach to what it termed a *portfolio approach*, as described in Box 2-1. The move toward a portfolio approach appears to add need-based and context-based flexibility to the IRIS program. EPA used chloroform as an example; it is developing a reference concentration for inhalation exposures and assessing whether the reference concentration protects against carcinogenic effects adequately. The decision to limit the assessment was based on consultation with EPA regulatory programs. Overall, the portfolio approach is expected to conserve agency resources, and it is consistent with the recommendations of the National Academies report, *Science and Decisions: Advancing Risk Assessment* (NRC 2009).

SYSTEMATIC REVIEW: PROBLEM FORMULATION, PROTOCOL DEVELOPMENT, AND EVIDENCE IDENTIFICATION AND EVALUATION

The 2014 report offered many recommendations related to systematic review, including problem formulation, protocol development, evidence identification, and evidence evaluation (Chapters 3–5, NRC 2014). The committee found that the IRIS program has made substantial progress in incorporating systematic-review methods into its process and assessments. Development and implementation of systematic-review methods have been facilitated by the recruitment of the current IRIS program director, who has extensive experience in the development of the methods and their application to chemical risk assessment. The IRIS program has also expanded internal training programs that are designed to improve staff understanding of the methods.

BOX 2-1 Environmental Protection Agency Description of Its Portfolio Approach

To ensure…support is timely and responsive, NCEA is developing a portfolio of chemical evaluation products employing the principles and state-of-the-art practices of systematic review. The portfolio approach will increase public health protection by:

- moving away from a "one-size-fits-all" approach to chemical risk assessment towards a spectrum of assessment products to meet specific decision contexts;
- facilitating the incorporation of new science into risk assessment and decision-making;
- tailoring assessments to meet the many needs of decision makers; and,
- increasing the number of chemicals that can be evaluated for their effects on human health by utilizing constrained resources in the most efficient manner.

Source: EPA (2018a).

Furthermore, the IRIS program has developed a number of formal and informal collaborations with groups that are active in systematic review, including the National Toxicology Program Office of Health Assessment and Translation, the World Health Organization (WHO), the European Food Safety Authority, the International Collaboration for Automation of Systematic Reviews, and the Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies (CAMARADES). Some of those collaborations help to position the IRIS program as a leader in advancing systematic-review methods, such as the development or modification of risk-of-bias tools for animal toxicity studies.

The committee was impressed by the efforts of IRIS program management to develop within the IRIS program the scientific expertise needed to conduct systematic reviews. Some notable changes have included the establishment of a systematic-review working group that should lead to increased efficiency and consistency among assessments. Other workgroups that are focused, for example, on epidemiology, physiologically based pharmacokinetic (PBPK) models, and neurotoxicology have been created; these teams of appropriate subject-matter experts are expected to support the IRIS process further through improved rigor of scoping and problem formulation and through improvements in other steps of the systematic-review process.

The 2014 report offered numerous recommendations related to systematic-review processes that are accepted as standards of practice in the scientific community. The present committee found multiple examples of the IRIS program's consideration and implementation of those recommendations, such as the development of systematic-review protocols, inclusion of an information specialist who is trained in systematic-review methods in the work groups, and the use of two-person teams for data extraction and risk-of-bias assessments. The IRIS program is also appropriately using a variety of software tools to assist with literature management (HERO), scoping (SWIFT), screening (Distiller), and data extraction (HAWC). The use of those and other software tools with input from appropriate subject-matter experts should improve efficiency, transparency, and rigor and directly address recommendations in the 2014 report. Many of the operational approaches used by the IRIS program are described in the assessment plans or the systematic-review protocols, and sufficient details are given to provide assurances that standardized systematic-review methods are being developed and applied by the IRIS program. The committee expects future systematic-review protocols to be streamlined and to become less generic when the handbook is completed.

The 2014 report also offered several recommendations about evaluating individual studies. Those recommendations encouraged EPA to use or develop tools for assessing risk of bias in different types of studies (human, animal, and mechanistic) and to add quality-assessment items relevant to particular systematic-review questions. EPA has implemented a process for evaluating risk of bias, and several documents that were provided to the committee (for example, EPA 2018b; Orem-Zavaleta 2018) demonstrate implementation of EPA's risk-of-bias tools and how EPA has augmented them with additional question-specific elements to assess study validity. The IRIS program, however, should provide information on the choice and use of tools, including its rationale for the choice of particular risk-of-bias domains. Including that documentation in the IRIS handbook will improve transparency. The committee notes that evaluation of risk of bias, although important, is not the only way to evaluate study quality. Accordingly, the IRIS program should show how other important methodologic characteristics of a particular study will be evaluated, and EPA should continue to seek and evaluate additional tools that can help to assess study quality.

As part of revisions of the IRIS process, EPA is producing assessment plans and systematic-review protocols. The committee found overlap between those documents; for example, PECO statements are found in both types of documents.[3] Indeed, the added value of a two-step process (assessment plan and protocol) was unclear to the committee. It was not immediately clear whether the assessment plan also serves as a "data call" for additional studies that are outside the scope specified by the systematic review but could inform the overall chemical-assessment process. Some additional clarification of terminology and clearer descriptions of how the documents will be used could help the public to understand how chemical assessments move through the IRIS process.

[3]A PECO statement is a structured framework that defines a question by specifying **p**opulation, **e**xposure, **c**omparator, and **o**utcome to be considered in a systematic review.

The committee identified several ways in which the IRIS program could benefit from refinements. For example, the link between scoping and problem formulation outlined in the assessment plan and development of the PECO statement was not well described. Improving the description of how scoping and problem formulation are used to focus the goals of a systematic review will lead to greater specificity in descriptions of outcomes, inclusion and exclusion criteria, and other elements found in the systematic-review protocol and will further improve the transparency and scientific rigor of the process. The committee found that the IRIS program included the dates and results of its literature searching and screening (for example, as appendixes) in draft systematic-review protocols that are undergoing public comment. Completing the literature search as part of protocol development is inconsistent with current best practices for systematic review, and the IRIS program is encouraged to complete the public-comment process and finalize the protocol before initiating the systematic review. Doing so will improve transparency in the IRIS process.

The committee identified several recommendations in the 2014 report that reflect broad scientific efforts that extend beyond the IRIS program. For example, several recommendations were related to the evaluation and use of mechanistic data in a systematic review. EPA's systematic-review process indicates that mechanistic data can be considered at various steps; for example, the draft protocol for the IRIS assessment of chloroform (EPA 2018b) describes how mechanistic data will be considered. Although appropriate tools, such as those to evaluate risk of bias in mechanistic studies, are in early stages of development in the broader scientific community, the IRIS program has developed approaches for the evaluation of PBPK models that will be used in assessments (Orme-Zavaleta 2018). The committee expects similar evaluation methods for other types of mechanistic evidence to emerge on a case-by-case basis and to include methods for determining at what stage and how mechanistic data could be used in an IRIS assessment. For example, mechanistic data were used by a National Academies committee to inform development of PECO statements for reproductive outcomes associated with *o*-phthalate compounds (NASEM 2017a). The committee notes that the use of mechanistic data by the IRIS program is consistent with other EPA programs, such as the Office of Pesticide Programs; for example, in the recent hazard identification conducted for benzo[a]pyrene (EPA 2017b), the IRIS program used mechanistic data extensively. Nonetheless, establishment of a framework for when and how mechanistic data would be identified, evaluated, and used remains challenging. The challenge is not unique to the IRIS program and has been identified for future work in the National Toxicology Program (NTP) handbook for conducting systematic reviews and evidence integration (NTP 2015a, p. 73–74).

Finally, the committee considered best practices for systematic reviews in other medical disciplines. Current best practices recommended by the Institute of Medicine (IOM 2011) suggest that the IRIS teams involved in the systematic-review process should be independent of those involved in regulatory decision-making who use the products of the systematic-review teams. The committee notes that the current organizational structure of the IRIS program in the EPA Office of Research and Development is consistent with those best practices.

EVIDENCE INTEGRATION

The 2011 report recommended standardizing an approach for synthesizing evidence within data streams (human, animal, and mechanistic) and integrating evidence across data streams (NRC 2011, p. 165).[4] From 2011 to 2013, the IRIS program moved solidly in that direction, as evidenced by its draft handbook (EPA 2013) and its example applications of the approach in two draft IRIS reports—the Toxicological Review of Methanol (Noncancer) and the Toxicological Review of Benzo[a]pyrene (see NRC 2014, pp. 93–96). Although the 2014 committee recognized that substantial progress had occurred during 2011–2013, it made several additional recommendations to guide the IRIS program toward a more systematic

[4]IRIS uses the phrase *evidence synthesis* to refer to the task of combining evidence from a given evidence stream, such as human or animal, and the phrase *evidence integration* to refer to the task of combining evidence from different evidence streams.

process for evidence synthesis and integration that would maximize transparency, efficiency, and scientific validity.

The major recommendation in Chapter 6 of the 2014 report guided IRIS to choose between making its current guided expert process more transparent and adopting a more structured, GRADE-like,[5] process along the lines of the NTP (NRC 2014, p. 105). The IRIS program has explicitly chosen the first option, using structured categories with criteria to guide expert judgment, and EPA has made substantial strides toward more systematic and transparent evidence synthesis (see slides 65–84, Appendix C; posters D-4 and D-5, Appendix D). Specifically, the IRIS program has created processes and guidelines for synthesizing human evidence and animal evidence that support choosing one category for characterizing the strength of evidence (see slides 82–84, Appendix C). The guidelines focus on human and animal evidence streams and use mechanistic evidence to inform evidence synthesis and to provide scientific guidance for evidence integration in the steps that follow. In using Bradford Hill criteria to move beyond association to causation and to build on the systematic evaluations of individual study quality conducted in the step before evidence synthesis,[6] the IRIS program has created a process for evidence synthesis that is scientifically consistent with the state of the art and that effectively leverages approaches of other programs, such as NTP, that face similar challenges. Increased transparency is evident in the examples and the workshop presentations, but further transparency would be obtained with completion of a handbook that provides more details about processes, reasoning behind decisions, and approaches for presenting results. In the interim, while EPA is completing its handbook, it is releasing protocols for each assessment that include a description of how evidence within each data stream will be synthesized and how evidence from multiple data streams will be integrated. The draft protocol for the IRIS assessment of chloroform (EPA 2018b) was provided as an example. The committee supports EPA's approach.

Integration of evidence across data streams was described by EPA in its presentations (see slides 79–87, Appendix C; posters D-4 and D-5, Appendix D) and in the draft chloroform protocol (EPA 2018b, pp. 43–53). Again, the process and framework within which evidence integration takes place (slides 82–84; Appendix C) are consistent with state-of-the-art approaches taken by other scientific institutions or agencies, such as NTP, that face similar challenges.

Some questions have been raised about the use of mechanistic data in evidence integration. When animal or human data are extensive, mechanistic data can be used to evaluate the evidence within or across the animal or human data streams rather than as a third stream of evidence to be analyzed separately and then combined with human and animal evidence. When extensive mechanistic data are available and human and animal data on apical end points are sparse, mechanistic data might be used reliably as a third data stream to identify hazards, as has been done for the dioxin-like polychlorinated biphenyls (IARC 2016). Mechanistic data are important in identifying potential adverse outcomes, including ones that are not evaluated in guideline-driven animal testing; in informing dose–response assessment; and in determining the relevance of animal data for human health risk estimation. For example, in the case of phthalates (poster D-7, Appendix D), mechanistic data were used to determine that not all effects on male reproductive development in rodents were relevant for humans, and the data provided a basis for selecting the studies that were most relevant as a starting point in establishing a reference dose. However, EPA acknowledged that understanding of mechanisms relevant to effects of phthalates on development is incomplete, and that uncertainty makes it difficult to estimate risk primarily on the basis of mechanistic information. Although organizing the body of evidence according to a mechanistic framework might at first seem desirable because of biologic relevance, mechanistic frameworks today could probably be completed for only a few chemicals. As noted in the 2014 report, solid conclusions about causality can be drawn without mechanistic information,[7] for example, when there is strong and consistent evidence from animal or epidemiology studies.

[5]GRADE is defined as grading of recommendations, assessment, development and evaluation.

[6]For example, see slide 69 in Appendix C, in which EPA advises using only medium-quality and high-quality studies and incorporating considerations of bias and sensitivity.

[7]"The history of science is replete with solid causal conclusions in advance of solid mechanistic understanding" (NRC 2014, p.90).

Another recommendation from Chapter 6 of the 2014 report concerns expanding EPA's capacity to perform quantitative evidence integration for hazard identification, for example, by using meta-regression or Bayesian analysis. To avoid compromising efficiency and timeliness in producing assessments, the 2014 report recommended developing such analytic capacities in parallel with other work in the IRIS program. EPA has taken the recommendation seriously and has explored meta-analytic approaches to combining animal data within species to determine whether the evidence indicates a chemical hazard, for example, whether trimethylbenzene poses a neurotoxic hazard (poster D-2, Appendix D). The IRIS program also initiated work on a Bayesian approach to combining data from different animal species (poster D-10, Appendix D).[8] The Bayesian work is promising, but application to IRIS assessments has not yet occurred. It is clear that the IRIS program has made progress here; the agency should continue with its efforts in this field.

Another recommendation from both the 2011 and the 2014 reports urged the use of more standardized, structured evidence tables to support the evidence-integration narrative[9] and emphasized the utility of a somewhat standard template for the narrative. The 2017 Toxicological Profile for Benzo[a]pyrene (EPA 2017b) provides an example of structured evidence tables that directly support the evidence-integration narrative, first for synthesis of individual data streams and then in an integrated summary form that connects evidential categorization with the supporting studies (Table 1-20, page 1-108). The final table lays out the evidence that the chemical is a human carcinogen by first introducing the human evidence on cancer from benzo[a]pyrene or precursors from complex mixtures and the human mechanistic studies and then discussing the findings of in vivo animal studies on tumors associated with multiple routes of exposure, adding the studies of precursor events, and finally presenting the evidence that precursor events are likely to occur in humans. The format is clear, well structured, and straightforward to follow. Although a well-reasoned discussion on noncancer effects is available in the same document, structured-narrative justifications of the evidence-integration process and the conclusion were not as well developed as those on cancer. In the workshop, EPA stated that standardized descriptors for noncancer effects are still needed and are being discussed within the agency.

EPA illustrated current thinking regarding the template for evidence integration in the workshop (slide 85, Appendix C) and in the chloroform draft protocol (EPA 2018b). The template has many characteristics of the GRADE approach to evaluating evidence, with similar labels for conclusions about the strength of the evidence within and across data streams. The approach appears to conform with the state of the art and bears considerable similarity to the system used by NTP (NTP 2015a,b). Although the chloroform protocol provides some illustration of EPA's approach, more detailed guidance and completed examples are needed to judge EPA's application of the template for evidence integration.

In summary, the IRIS program has made substantial strides in meeting the recommendations of the 2011 and 2014 reports regarding synthesis and integration of evidence. The IRIS process that was presented to the committee is dramatically more systematic, transparent, and scientifically defensible than the one presented in the 2010 IRIS Toxicological Review of Formaldehyde (EPA 2010).

DERIVATION OF TOXICITY VALUES

Recommendations regarding derivation of toxicity values were provided in Chapter 7 of the 2014 report. An important recommendation in that chapter was to "develop criteria for determining when evidence is sufficient to derive toxicity values." In the workshop, EPA described the overall process and criteria that the agency intends to use to implement that recommendation and indicated that it would develop toxicity values when the evidence-integration conclusion is the "strongest" or a "moderately strong conclusion for a human health effect." As noted, EPA clarified that the agency intends to systematize processes

[8]The Bayesian approach is based on the seminal work of Dumouchel and Harris (1983) and recent work of Jones et. al. (2009).

[9]An evidence-integration narrative is a description of the available evidence and the argument for or against a particular hazard.

to maintain transparency in reaching the hazard conclusion (slides 132–133, Appendix C), although standard descriptors for noncancer effects are being reviewed within the agency and are not yet final.

EPA's approach is consistent with the 2014 recommendation that formal dose–response assessments should be restricted to outcomes on which evidence integration has led to the strongest or a moderately strong conclusion on the given health effect, such as known or likely to be carcinogenic to humans (slide 131, Appendix C). EPA indicated that when there is less strong evidence on a human health effect, such as suggestive evidence of cancer, the decision to develop a toxicity value will be determined by the situation (for example, when there is a well-conducted study and a value would be useful for a decision). However, EPA did not present criteria to be used in such cases.

The one example in which criteria have been applied to support the derivation of toxicity values was the chloroprene reassessment (Orme-Zavaleta 2018). In that document, EPA focused its systematic review on publications since the 2010 assessment. EPA concluded that the new studies did not change the conclusions in the 2010 assessment and did not justify a reassessment of human health effects (that is, derivation of new toxicity values). Although commenting on the conclusions in that assessment is beyond the scope of the present committee's task, the committee acknowledges that EPA's reassessment described its criteria for evaluating risk of bias and study sensitivity needed to detect a true effect and that it presented criteria for evaluating PK/PBPK studies. Furthermore, EPA explained why each study considered in the final assessment did not change the conclusions reached in the 2010 IRIS assessment and did not justify a reassessment of human health effects. Thus, it is clear that EPA is making progress toward improving transparency in its use of systematic review and expert judgment to inform the derivation of toxicity values directly.

Another important recommendation in the 2014 report was that EPA "continue its shift toward the use of multiple studies rather than single studies for dose–response assessment" (NRC 2014). The present committee noted that progress has been made in the use of multiple studies for dose–response assessments as exemplified in the recent assessments of ethylene oxide and benzo[a]pyrene (slides 134–135, Appendix C) and builds on efforts to compare candidate reference doses or concentrations in previous assessments, such as in the 2012 IRIS Toxicological Review of Tetrachloroethylene (EPA 2012). EPA is further developing new tools for visualizing comparisons to communicate the outcome of assessments more effectively, as was demonstrated in the workshop by using HAWC. EPA acknowledged, and the committee agrees, that the development of systematic assessments for many types of mechanistic studies that could contribute to the assessment remains challenging, not only to EPA but to the scientific community generally. However, the process that EPA previously developed to review PK/PBPK models and to describe how they could be used in dose–response and toxicity-value assessments (EPA 2006) is a good example of best practices. As other forms of mechanistic data become more readily available, partly driven by previous National Academies reports (NASEM 2017b; NRC 2007), the IRIS program should develop new approaches for using such studies to inform dose–response and toxicity-value assessments (slides 142–147, Appendix C). Such guidance will improve transparency and encourage new science, whether it is used to support evidence of potential toxicity or, just as important, to provide perspectives on the potential exposure conditions that could reasonably be expected to cause toxicity.

The 2014 report also recommended that EPA use formal methods for combining multiple studies and further develop and expand its use of Bayesian and other formal quantitative methods for dose–response assessment and derivation of toxicity values (NRC 2014). EPA has begun to develop and apply tools for meta-regression analysis and Bayesian approaches and has demonstrated their application in case studies (slides 135, 136, 139, and 140, Appendix C; posters D-2 and D-10, Appendix D). Implementation of the recommendation will continue and will require sustained resources and continued capacity-building to develop a process that is ultimately transparent, is replicable, and represents best practices for the future. And it will require close collaborations between domain experts in the biologic and mathematical or statistical disciplines within EPA and with other agencies and stakeholders to establish clear criteria and guidance, including articulation of underlying assumptions, strengths, and weaknesses of each approach. The committee notes that care must be taken when combining results within or between studies in developing dose–response relationships inasmuch as multiple mechanisms, each with its own potential dose–response relationship, might be involved. In such cases, clearly articulated expert judgment, criteria for expert selection,

and multidisciplinary collaborations need to be supported and used in the development and application of new mathematical approaches.

The 2014 report recommended that EPA develop IRIS-specific guidelines to frame analysis and communication of uncertainty (NRC 2014). EPA has made substantial progress in developing and adopting tools to address uncertainty analysis and communication (slides 136–138, Appendix C; poster D-6, Appendix D). It demonstrated its work during the workshop and focused on model uncertainty (slide 136, Appendix C) and the probabilistic distribution of toxicity values (slides 137–138, Appendix C). It further indicated that the IRIS program intends to adopt the WHO/International Programme on Chemical Safety guidance (slide 137, Appendix C) and to provide various calculations when reporting toxicity values, including ranges of risk-specific toxicity values (slide 138, Appendix C) to demonstrate uncertainty. The committee recognizes that the steps taken are in the right direction for an evolving process and encourages EPA to continue to develop and test new tools in collaboration with other agencies and stakeholders. Equally important, the committee encourages EPA to continue its effort to frame uncertainty analysis and communications to address multiple sources of uncertainty surrounding toxicity values.

CONCLUDING REMARKS

Overall, the committee is encouraged by the steps that EPA has taken, which have accelerated during the last year under new leadership. During the workshop, the committee was impressed by the overall enthusiasm displayed by EPA staff and the substantive progress toward full implementation of systematic review and transparency in IRIS assessments. The committee fully appreciates that changing the process and implementing systematic-review procedures while producing final assessments is a huge challenge for any organization, especially in such a short period (12 months) and with a shrinking staff. Because new tools and approaches will ultimately be needed to implement past National Academies recommendations, especially for incorporating mechanistic information and for integrating evidence across studies, the committee is encouraged by IRIS program efforts to collaborate with other EPA staff, other government agencies, and academe to have the right mix of expertise to develop new approaches and best practices for conducting assessments.

REFERENCES

DuMouchel, W.H., and J.E. Harris. 1983. Bayes methods for combining the results of cancer studies in humans and other species: Rejoinder. J. Am. Stat. Assoc. 78(382):313-315.

EPA (US Environmental Protection Agency). 2006. Approaches for the Application of Physiologically Based Pharmacokinetic (PBPK) Models and Supporting Data in Risk Assessment. EPA/600/R-05/043F. National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC. August 2006. Available: https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=157668 [accessed March 6, 2018].

EPA. 2010. IRIS Toxicological Review of Formaldehyde (Interagency Science Consultation Draft). EPA/635/R-10/002C. U.S. Environmental Protection Agency, Washington, DC [online]. Available: https://cfpub.epa.gov/si/si_public_record_report.cfm?direntryid=223603 [accessed February 27, 2018].

EPA. 2012. Toxicological Review of Tetrachloroethylene (Perchloroethylene) [CAS No. 127-18-4]. EPA/635/R-80/011F. U.S. Environmental Protection Agency, Washington, DC. February 2012 [online]. Available: https://cfpub.epa.gov/ncea/iris/iris_documents/documents/toxreviews/0106tr.pdf [accessed February 27, 2018].

EPA. 2013. Part I: Status of Implementation of Recommendations. Materials Submitted to the National Research Council. Integrated Risk Information System Program, U.S. Environmental Protection Agency. January 30, 2013 [online]. Available: https://www.epa.gov/sites/production/files/2014-06/documents/iris_program_materials_to_nrc_part_1.pdf [accessed February 23, 2018].

EPA. 2017a. IRIS Assessment Plan for Chloroform [CASRN 67-66-3]. EPA/635/R-17/330. Integrated Risk Information System, National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC.

EPA. 2017b. Toxicological Review of Benzo[a]pyrene. EPA/635/R-17/003Fa. Integrated Risk Information System, National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC [online]. Available: https://cfpub.epa.gov/ncea/iris/iris_documents/documents/toxreviews/0136tr.pdf [accessed February 14, 2018].

EPA. 2018a. A Message from the IRIS Program, January 2018. IRIS Recent Additions, January 25, 2018 [online]. Available: https://www.epa.gov/iris/iris-recent-additions [accessed February 14, 2018].

EPA. 2018b. Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation) [CASRN 67-66-3]. EPA/635/R-17/486. Integrated Risk Information System, National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC.

EPA. 2018c. IRIS Assessment Plan for Uranium (Oral Reference Dose) (Scoping and Problem Formulation Materials) [CASRN 7440-61-1]. EPA/635/R-17/787. Integrated Risk Information System, National Center for Environmental Assessment, Office of Research and Development.

IARC (International Agency for Research on Cancer). 2015. Polychlorinated Biphenyls and Polybrominated Biphenyls. IAEC Monographs on the Evaluation of Carcinogenic Risks to Humans Vol. 107 [online]. Available: http://monographs.iarc.fr/ENG/Monographs/vol107/mono107.pdf [accessed March 22, 2018].

IOM (Institute of Medicine). 2011. Finding What Works in Health Care: Standards for Systematic Review. Washington, DC: The National Academies Press.

Jones, D.R., J. Peters., J.L. Rushton, A.J. Sutton, and K.R. Abrams. 2009. Interspecies extrapolation in environmental exposure standard setting: A Bayesian synthesis approach. Regul. Toxicol. Pharmacol. 53(3):217-225.

National Academies of Sciences, Engineering, and Medicine. 2017a. Application of Systematic Review Methods in an Overall Strategy for Evaluating Low-Dose Toxicity from Endocrine Active Chemicals. Washington, DC: National Academies Press.

National Academies of Sciences, Engineering, and Medicine. 2017b. Using 21st Century Science to Improve Risk-Related Evaluations. Washington, DC: National Academies Press.

NRC (National Research Council). 2007. Toxicity Testing in the 21st Century: A Vision and a Strategy. Washington, DC: The National Academies Press.

NRC. 2009. Science and Decisions: Advancing Risk Assessment. Washington, DC: The National Academies Press.

NRC. 2011. Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde. Washington, DC: The National Academies Press.

NRC. 2014. Review of EPA's Integrated Risk Information System (IRIS) Process. Washington, DC: The National Academies Press.

NTP (National Toxicology Program). 2015a. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Office of Health Assessment and Translation (OHAT), Division of the National Toxicology Program, National Institute of Environmental Health Sciences, U.S. Department of Health and Human Services, January 9. Available: https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf [accessed March 6, 2018].

NTP. 2015b. Handbook for Preparing Report on Carcinogens Monographs. Office of Health Assessment and Translation (OHAT), Division of the National Toxicology Program, National Institute of Environmental Health Sciences, U.S. Department of Health and Human Services, July 20, 2015. Available: https://ntp.niehs.nih.gov/ntp/roc/handbook/roc_handbook_508.pdf [accessed March 6, 2018].

Orme-Zavaleta, J. 2018. Response to the Request for Correction (RFC). Letter to Robert Holden, Liskow & Lewis, New Orleans, LA, from Jennifer Orme-Zavaleta, Principal Deputy Assistant Administrator for Science, Office of Research and Development, Washington, DC, January 25, 2018; Attachment 1. EPA Response to the Denka Performance Elastomers (DPE) Request for Correction of the Toxicological Review of Chloroprene (CAS No. 126-99-8) In Support of Summary Information on the Integrated Risk Information System (IRIS); Attachment 2. Systematic Review of Chloroprene [CASRN 126-99-80] Studies Published Since 2010 IRIS Assessment to Support Consideration of the Denka Request for Correction (RFC). January 2018 [online]. Available: https://www.epa.gov/sites/production/files/2018-01/documents/epa_repsonse_to_mr._holdren_jan_25_2018_complete.pdf [accessed February 9, 2018].

Appendix A

Biographic Information on the Committee to Review Advances Made to the IRIS Process

Jonathan M. Samet (*Chair*) is a pulmonary physician and epidemiologist. He is the dean of the Colorado School of Public Health and previously served as a professor and Flora L. Thornton Chair of the Department of Preventive Medicine of the Keck School of Medicine of the University of Southern California (USC) and director of the USC Institute for Global Health. His research has focused on the health risks posed by inhaled pollutants. He has served on numerous committees concerned with public health: the US Environmental Protection Agency Science Advisory Board; committees of the National Academies, including chairing the Biological Effects of Ionizing Radiation VI Committee, the Committee on Research Priorities for Airborne Particulate Matter, the Committee to Review EPA's Draft IRIS Assessment of Formaldehyde, the Committee to Review the IRIS Process, and the Board on Environmental Studies and Toxicology; and the National Cancer Advisory Board. He is a member of the National Academy of Medicine. Dr. Samet received his MD from the University of Rochester School of Medicine and Dentistry.

Richard A. Corley (retired) was a laboratory fellow at the Pacific Northwest National Laboratory operated by Battelle for the US Department of Energy. Dr. Corley specializes in the development of physiologically based pharmacokinetic models, real-time breath analysis, dermal and inhalation bioavailability, and the development of three-dimensional computational fluid-dynamic models of the respiratory system. He has published numerous peer-reviewed papers on oral, dermal, and inhalation toxicology; on modes of action of a variety of industrial and consumer chemicals; and on pharmacokinetic modeling and its applications in human health risk assessment. Dr. Corley has served on several National Academies committees, including the Committee to Assess the Health Implications of Perchlorate Ingestion, the Standing Committee on Risk Analysis Issues and Reviews, the Committee to Review EPA's Draft IRIS Assessment of Formaldehyde , and the Committee to Review EPA's Draft State of the Science Paper on Nonmonotonic Dose Response. He received a PhD in environmental toxicology from the University of Illinois at Urbana-Champaign.

George Daston is the Victor Mills Society Research Fellow at the Procter & Gamble Company. He has published over 100 articles and book chapters and edited five books in toxicology and risk assessment. His current research efforts are in toxicogenomics and mechanistic toxicology, particularly addressing how findings in these fields can improve risk assessment of chemicals and the development of nonanimal alternatives. Dr. Daston has served as president of the Teratology Society, as councilor and treasurer-elect of the Society of Toxicology, and on the US Environmental Protection Agency Science Advisory Board, the Board of Scientific Counselors of the National Toxicology Program, the National Academies Board on Environmental Studies and Toxicology, and the National Children's Study Advisory Committee. He is editor-in-chief of *Birth Defects Research: Developmental and Reproductive Toxicology*. With scientists at the US Humane Society, Dr. Daston manages the AltTox Web site, which is devoted to the exchange of scientific information leading to the development of in vitro replacements for toxicity assessments. Dr. Daston has been awarded the Josef Warkany Lectureship and the Distinguished Service Award by the Teratology Society, the George H. Scott Award by the Toxicology Forum, and the Society of Toxicology's Best Paper of the Year Award, and he is an elected fellow of the American Association for the Ad-

vancement of Science. Dr. Daston is an adjunct professor of pediatrics at the University of Cincinnati. He earned his PhD in developmental biology from the University of Miami.

David Dorman is a professor of toxicology in the Department of Molecular Biomedical Sciences at North Carolina State University. His research interests include neurotoxicology, nasal toxicology, pharmacokinetics, and cognition and olfaction in animals. He has served on advisory boards for the US Navy, the National Aeronautics and Space Administration, the US Department of Agriculture, and the National Toxicology Program. He has chaired several National Academies committees, including the Committee on Endocrine-Related Low Dose Toxicity, the Committee on Predictive-Toxicology Approaches for Military Assessments of Acute Exposures, and the Committee on Design and Evaluation of Safer Chemical Substitutions. He was also a member of the Committee to Review EPA's Draft IRIS Assessment of Formaldehyde and the Committee to Review the IRIS Process. Dr. Dorman is an elected fellow of the Academy of Toxicological Sciences, is a fellow of the American Association for the Advancement of Science, and is a national associate of the National Academies of Sciences, Engineering, and Medicine. He received a DVM from Colorado State University and completed a combined PhD and veterinary toxicology residency program at the University of Illinois at Urbana-Champaign. Dr. Dorman is a diplomate of the American Board of Veterinary Toxicology and the American Board of Toxicology.

Russ Hauser is the chair of the Department of Environmental Health, Frederick Lee Hisaw Professor of Reproductive Physiology, and professor of environmental and occupational epidemiology at the Harvard T.H. Chan School of Public Health. He also holds an appointment at the Harvard Medical School, where he is professor of obstetrics, gynecology, and reproductive biology. Dr. Hauser's research focuses on the effects of environmental chemicals on reproductive health, pregnancy, and children's health. He has served on several National Academies committees, including the Committee to Review EPA's State of the Science Paper on Nonmonotonic Dose Response, the Committee on the Health Risks of Phthalates, and the Committee on Endocrine-Related Low-Dose Toxicity. Dr. Hauser was a member of two US Environmental Protection Agency Science Advisory Boards, served on the US Consumer Product Safety Commission's Chronic Hazard Advisory Panel that examined the effects of phthalates on children's health, and is an associate editor of *Environmental Health Perspectives*. He received his MD from the Albert Einstein College of Medicine and his MPH and ScD from the Harvard T.H. Chan School of Public Health, where he also completed a residency in occupational medicine. He is board-certified in occupational medicine.

Karen A. Robinson is an associate professor at the Johns Hopkins University School of Medicine. She also serves as director of the Johns Hopkins University Evidence-Based Practice Center and is a member of the core faculty in the Center for Clinical Trials and Evidence Synthesis at the university's Bloomberg School of Public Health. Her research focuses on evidence-based health care and evidence-based research. She conducts systematic reviews that are used to develop clinical practice guidelines and to inform other health decisions. She served on the National Academies Committee on Endocrine-Related Low-Dose Toxicity and Committee on Gulf War and Health: Treatment of Chronic Multisymptom Illness. Dr. Robinson received her MSc in health sciences from the University of Waterloo, Ontario, and her PhD in epidemiology from the Bloomberg School of Public Health.

Richard P. Scheines is a professor of philosophy and dean of the Dietrich College of Humanities and Social Sciences of Carnegie Mellon University. His research focuses on causal discovery, specifically the problem of learning about causation from statistical evidence. Dr. Scheines also works in building and researching the effectiveness of educational software, including intelligent proof tutors and virtual causality laboratories, and a full-semester course on causal and statistical reasoning. Because of that work, he has a courtesy appointment in the Human-Computer Interaction Institute of Carnegie Mellon. He served on several National Academies committees, including the Committee to Review the IRIS Process. Dr. Scheines received a PhD in the history and philosophy of science from the University of Pittsburgh.

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

Lauren Zeise is director of the California Environmental Protection Agency's Office of Environmental Health Hazard Assessment. She oversees the department's activities, which include the development of risk assessments, hazard evaluations, toxicity reviews, cumulative impact analyses, frameworks and methods for assessing toxicity and cumulative effects of vulnerability and environmental exposures on communities, and the department's activities in the California Environmental Contaminant Biomonitoring Program. Dr. Zeise was the 2008 recipient of the Society for Risk Analysis Outstanding Practitioners Award. She has served on advisory boards and committees of the US Environmental Protection Agency, the Office of Technology Assessment, the World Health Organization, and the National Institute of Environmental Health Sciences. Dr. Zeise has served on numerous National Academies committees, including the Committee on Toxicity Testing and Assessment of Environmental Agents and the Committee on Improving Risk Analysis Approaches Used by the U.S. Environmental Protection Agency. Dr. Zeise received a PhD from Harvard University.

Yiliang Zhu is a professor in the Division of Epidemiology, Biostatistics, and Preventive Medicine of the University of New Mexico (UNM) School of Medicine. He directs the biostatistics, epidemiology, and research design cores for the UNM Clinical and Translational Research Center and for the Mountain West Clinical and Translational Research Infrastructure Network, a consortium of 13 universities in seven states. His research focuses on quantitative methods in health risk assessment, including integrative modeling of biologic systems, dose–response modeling, benchmark-dose methods, and uncertainty quantification. He also conducts research in biostatistics methods, clinical- and health-outcome evaluation, and impact assessment of health-care systems and policies in northwestern rural China. Before joining UNM, Dr. Zhu was a professor at the University of South Florida College of Public Health where he directed the Biostatistics PhD program and the Center for Collaborative Research. Dr. Zhu has served on several National Academies committees, including the Committee on EPA's Exposure and Human Health Assessment of Dioxin and Related Compounds, the Committee on Tetrachloroethylene, the Committee to Review EPA's Draft IRIS Assessment of Formaldehyde, and the Committee to Review the IRIS Process. He received a PhD in statistics from the University of Toronto.

Appendix B

Open Session Workshop Agenda

COMMITTEE TO REVIEW ADVANCES MADE TO THE IRIS PROCESS

SECOND MEETING

Open Session: February 1-2, 2018
National Academies of Sciences, Lecture Room
2101 Constitution Ave, NW
Washington, DC 20418

OPEN SESSION AGENDA

9:30 **Purpose of Open Session and Introduction of Committee Members**
Jonathan Samet
Chair, Committee to Review Advances Made to the IRIS Process
Dean, Colorado School of Public Health

9:45 **Introduction and Overview of Improvements to the IRIS Program**
Tina Bahadori
Director, National Center for Environmental Assessment
U.S. Environmental Protection Agency

Kristina Thayer
Director, Integrated Risk Information System (IRIS) Division
U.S. Environmental Protection Agency

10:45 **Discussion with National Academies Committee**

11:30 **Opportunity for Public Comments to National Academies Committee**

12:00 *Lunch Break*

1:00 **Session 1: Systematic Review in the IRIS Program – Evidence Identification**

EPA Panel Presentations and Discussion with the National Academies Committee on the Following Topics:

Scoping, Problem Formulation, and Protocols
Literature Searching, Screening, and Inventories

2:00 **Opportunity for Public Comments to National Academies Committee**

2:15 **Session 2: Systematic Review in the IRIS Program – Evidence Evaluation**

EPA Panel Presentations and Discussion with the National Academies Committee on the Following Topics:

Evaluating Individual Studies: Reporting Quality, Risk of Bias, and Sensitivity

Evaluating Confidence in a Body of Evidence: Evidence Synthesis and Integration to Reach Hazard Conclusions

3:15 **Opportunity for Public Comments to National Academies Committee**

3:30 *Break*

3:45 **Session 3: Development and Application of Specialized Tools for Systematic Review**

EPA Panel Presentations and Discussion with the National Academies Committee

4:30 **Opportunity for Public Comments to National Academies Committee**

5:00 *Break*

5:30-7:00 **Poster Session and Demonstrations, West Court**

FRIDAY, FEBRUARY 2, 2018

8:30 **Welcome and Recap from First Day**
Jonathan Samet
Chair, Committee to Review Advances Made to the IRIS Process
Dean, Colorado School of Public Health

8:45 **Session 4: Study Selection for Developing Toxicity Values, and Advancing Research on Quantitative Analyses for Evidence Integration and Dose-Response Analyses**

EPA Panel Presentations and Discussion with the National Academies Committee

10:00 **Opportunity for Public Com ments to National Academies Committee**

10:15 *Break*

10:30 **Collaborations, Training, and Final Thoughts**
Tina Bahadori
Director, National Center for Environmental Assessment
U.S. Environmental Protection Agency

Kristina Thayer
Director, Integrated Risk Information System (IRIS) Division
U.S. Environmental Protection Agency

11:00 **Discussion with National Academies Committee**

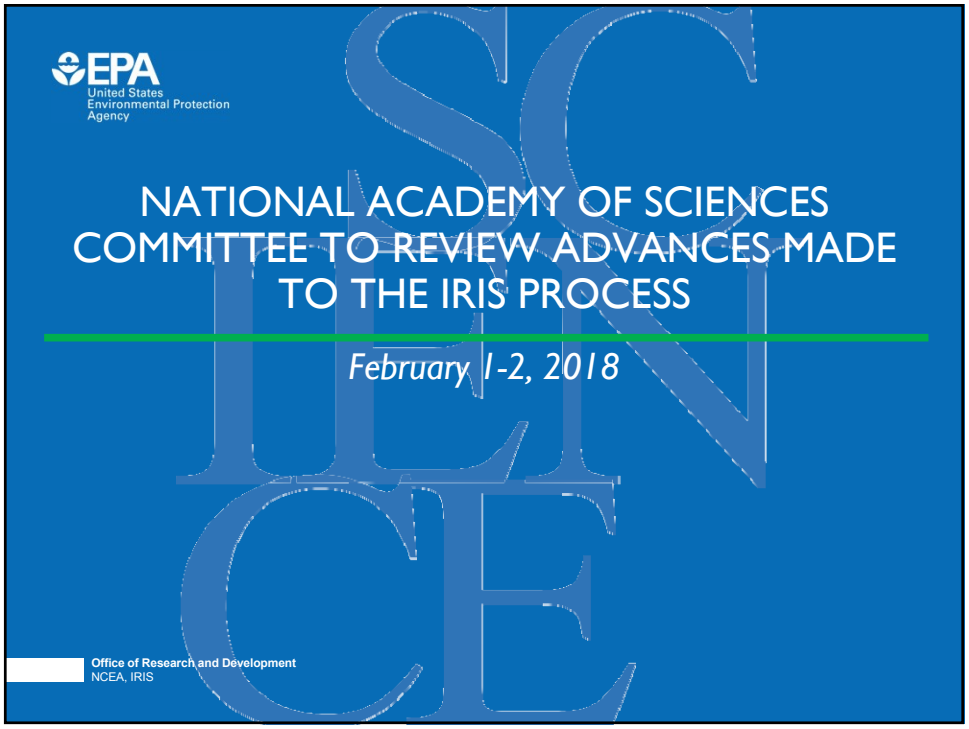
11:45 **Opportunity for Public Comments to National Academies Committee**

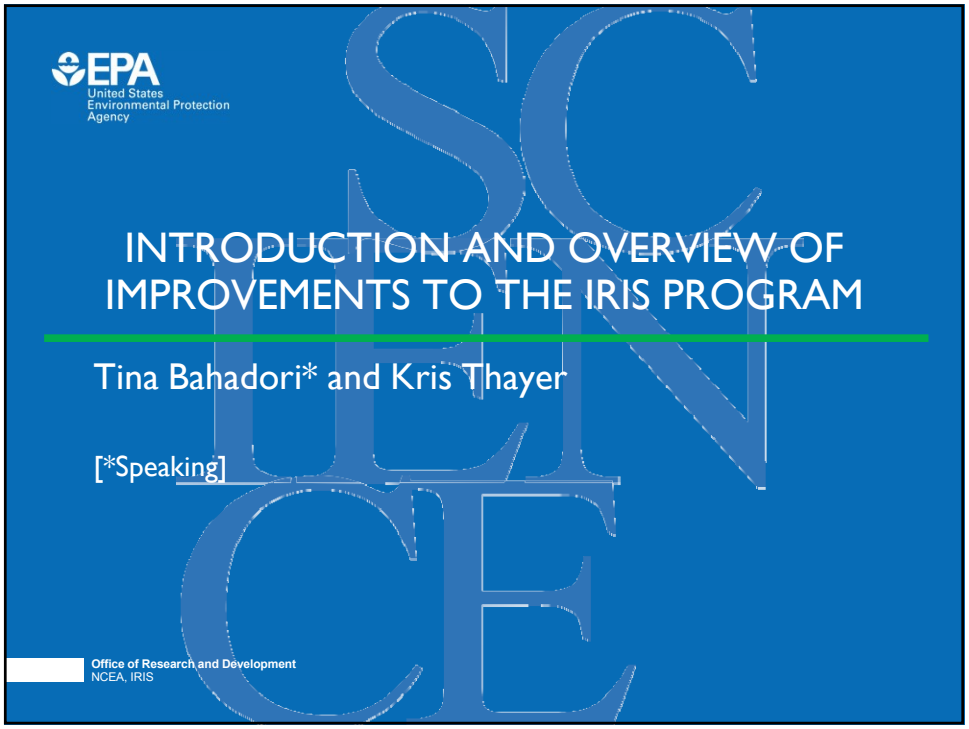
12:30 *Adjourn*

Appendix C

Presentations by US Environmental Protection Agency

Appendix C





Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

EPA IRIS

- **Created in 1985 to foster consistency in the evaluation of chemical toxicity across the Agency.**
- **IRIS assessments contribute to decisions across EPA and other health agencies.**
- **Toxicity values**
  - Noncancer: Reference Doses (RfDs) and Reference Concentrations (RfCs).
  - Cancer: Oral Slope Factors (OSFs) and Inhalation Unit Risks (IURs).
- **IRIS assessments have no direct regulatory impact until they are combined with**
  - Extent of exposure to people, cost of cleanup, available technology, etc.
  - Regulatory options.
  - Both of these are the purview of EPA's program offices.

2

EPA

IRIS Provides Scientific Foundation for Agency Decision Making

IRIS

- **Clean Air Act (CAA)**
- **Safe Drinking Water Act (SDWA)**
- **Food Quality Protection Act (FQPA)**
- **Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA)**
- **Resource Conservation and Recovery Act (RCRA)**
- **Toxic Substances Control Act (TSCA)**

Broad Input to Support →
- **Agency Strategic Goals**
- **Children's Health**
- **Environmental Justice**

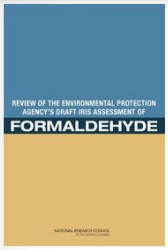
3

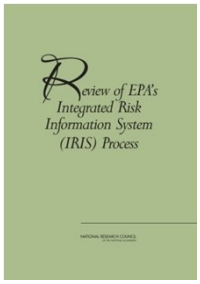
New Leadership Structure in NCEA

- **In January 2017, EPA appointed new leadership to the National Center for Environmental Assessment and to its IRIS Program.**
  - **NCEA Director:** significant experience in the chemical and energy industries, and formerly the Director of ORD's Chemical Safety for Sustainability National Research Program, Tina Bahadori brings knowledge of TSCA, innovative applications of computational toxicology, and exposure science.
  - **IRIS Program Director:** As a recognized leader in systematic review, automation, and chemical evaluations, Kris Thayer brings experience in early partner and stakeholder engagement and input, and demonstrated actions to increase capacity and transparency in assessments.
- **Improved responsiveness and accountability through Senior Leadership Team.**
- **Integrating across the spectrum of human and ecological RA practices.**

4

Drivers for this Study





GAO
Accountability * Integrity * Reliability

https://www.gao.gov/highrisk/transforming_epa_and_toxic_chemicals/why_did_study

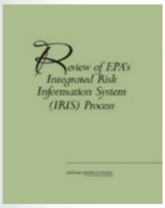
Fiscal Year 2017 Appropriations
https://www.congress.gov/114/crpt/srpt281/CRPT-114srpt281.pdf

5

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

NAS (2014) Overarching Statements

2014



"Overall, the committee finds that substantial improvements in the IRIS process have been made, and it is clear that EPA has embraced and is acting on the recommendations in the NRC formaldehyde report. The NRC formaldehyde committee recognized that its suggested changes would take several years and an extensive effort by EPA staff to implement. Substantial progress, however, has been made in a short time..." [p.9]

"EPA has not only responded to the recommendations made in the NRC formaldehyde report but is well on the way to meeting the general systematic-review standards for identifying and assessing evidence." [p. 51]

"... the IRIS program has moved forward steadily in planning for and implementing changes in each element of the assessment process. The committee is confident that there is an institutional commitment to completing the revisions of the process..." [p.135]

"The committee commends EPA for its substantive new approaches, continuing commitment to improving the process, and successes to date. Overall the committee expects that EPA will complete its planned revisions in a timely way and that the revisions will transform the IRIS Program." [p.135]

6

Previous Phased Improvements to the IRIS Program

- **Revising the structure of assessments to enhance the clarity and transparency of presentation:**
  - Detailing the methods underlying each step of draft development (e.g., literature search strategy).
  - Restructuring the document into separate hazard identification and dose-response chapters.
  - Replacing lengthy study summaries with synthesis text, supported by standardized tables and graphs.
- **Implementing "IRIS Enhancements"**
  - An updated process for developing and reviewing assessments that increases public input and peer consultation at earlier stages of assessment development, and clarifies processes for considering new evidence and scientific issues.
- **Establishing the SAB Chemical Assessment Advisory Committee (CAAC)**
  - 5 IRIS assessments completed CAAC review since 2014.
- **Restructuring the IRIS Program to create expertise-specific workgroups and improved assessment oversight.**

7

Quality Management

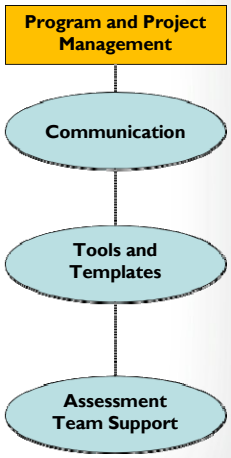
- **Assessment Development and Review**
  - Quality management inherent to systematic review methodology (e.g., independent screening of studies)
  - Rigorous review process includes internal, public, and external peer review
- **Scientific Support Teams**
  - Systematic review methods (Systematic Review Workgroup)
  - Systematic review support to chemical assessment teams (e.g., screening, study evaluation, data extraction, use of specialized software, etc. – train the trainer model)
  - Discipline-specific workgroups (e.g., epidemiology, PBPK, neurotoxicology, etc.)
  - Executive oversight
- **Roles and Responsibilities**
  - Assessment plans, protocols, and draft assessments indicate contributors and roles
  - Given current budget there is very limited use of contract support to conduct assessments
- **Training**
  - regular training via skill-building seminars, focused discussions, and retreats

8

Improved Practices for Timeliness and Resource Management

Current Program and Project Management in IRIS:

- *Centralized communication processes* for providing staff with updates on near-term priorities, template materials, and other process-oriented decisions.
- *Development and maintenance* of templates and checklists for key steps of assessment development using Microsoft SharePoint and Project as collaborative, web-based tools for assessment teams and project managers (document management and storage; scheduling support).
- *Dedicated IRIS Program staff and on-site programmatic contractor support* to facilitate continued implementation of program and project management principles.



9

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

GAO 2017 Report

- **Acknowledged the actions ORD has taken to enable the IRIS Program to produce timely, transparent, and credible assessments in support of EPA's mission.**
- **Discussions with GAO during and after the release of the 2017 High Risk Report have focused on approaches to demonstrate how management and integrity initiatives within IRIS are supporting the transformation of the program**

Summary of 2015 and 20	7 GAO High Risk Criteria Rat	ings of the IRIS Program
GAO High Risk Criteria	2015 Rating	2017 Rating
Leadership Commitment	Met	Met
Monitoring	Partially Met	Met
Action Plan	Partially Met	Partially Met
Demonstrated Progress	Not Met	Partially Met
Capacity	Not Met	Partially Met

- **IRIS is engaged in continual ongoing discussion with GAO regarding recommendations from the 2008, 2012, and 2013 reports.**
- **Of the seventeen recommendations issued in these three reports, as of June 2017, we have successfully closed ten recommendations and are rapidly moving to address the remaining seven.**

10

IRIS Multi-Year Agenda

- **Released to the public December 2015**
  - Result of a survey EPA program and regional offices for their assessment needs balanced with resource availability.
  - Other chemicals were also carried over from earlier prioritizations
  - Reflects global priorities
- **In FY 2018, reaffirm priorities; identify new or more urgent needs.**
- **Engage states.**

Group	Chemicals
1	Manganese
	Mercury/methylmercury
	Nitrate/nitrite
	Perfluoroalkyl compounds
	Vanadium and compounds
2	Acetaldehyde
	Ammonia (oral)
	Cadmium and compounds
	Uranium
3	Di-(2-ethylhexyl) phthalate
	Dichlorobenzene isomers
	Methyl t-butyl ether (MTBE)
	Nickel and compounds
	Styrene

11

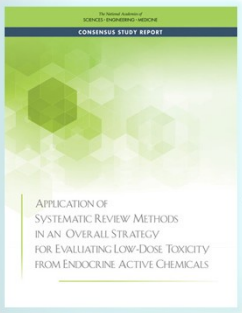
A Portfolio Approach

- **Moving away from a 'one-size-fits-all' approach to risk assessment towards a spectrum of assessment products to meet specific decision contexts;**
- **Facilitating the incorporation of new science into risk assessment and decision-making;**
- **Enabling assessments to be better tailored to meet needs of decision makers;**
- **Increasing the number of chemicals that can be evaluated for their effects on human health by utilizing constrained resources in the most efficient manner.**

12

Leading Edge of Science – Systematic Review

NAS 2017: Reflections and Lessons Learned from the Systematic Review



"….one disadvantage in conducting a systematic review is that it can be time and resource intensive, particularly for individuals that have not previously conducted a systematic review." [p.157]

"The committee discussed at length whether it could provide EPA with advice about when a systematic review should be performed but decided it could not be more specific because that decision will depend on the availability of data and resources, the anticipated actions, the time frame for decision making, and other factors." [p.157]

" committee also recognized that it might be advantageous for EPA to build on existing systematic reviews that are published in the peer-reviewed literature." [p.157]

" committee recognizes that the methods and role of systematic review and meta-analysis in toxicology are evolving rapidly and EPA will need to stay abreast of these developments, strive for transparency, and use appropriate methods to address its questions." [p.157]

13

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

Leading Edge of Science – New Data Streams

Next Generation IRIS

- IRIS in the 21st Century – implement recommendations of the NAS 2017 report, Using 21st Century Science to Improve Risk-Related Evaluations;
- New Approach Methods – see poster session
- Collaborate with Tox21
  - build expert-judgment case studies that inform assessment development and fill gaps in assessments, especially for data poor chemicals;
  - inform where resources should be strategically invested to generate additional data.
- Create efficiencies – engage other agencies to share common practices, data, and tools, and more efficiently leverage resources across the federal government.
- Refresh science – MOU's with academia and other federal agencies; strategic staffing; deeper engagement with health agencies in states.

14

How is IRIS Evolving?

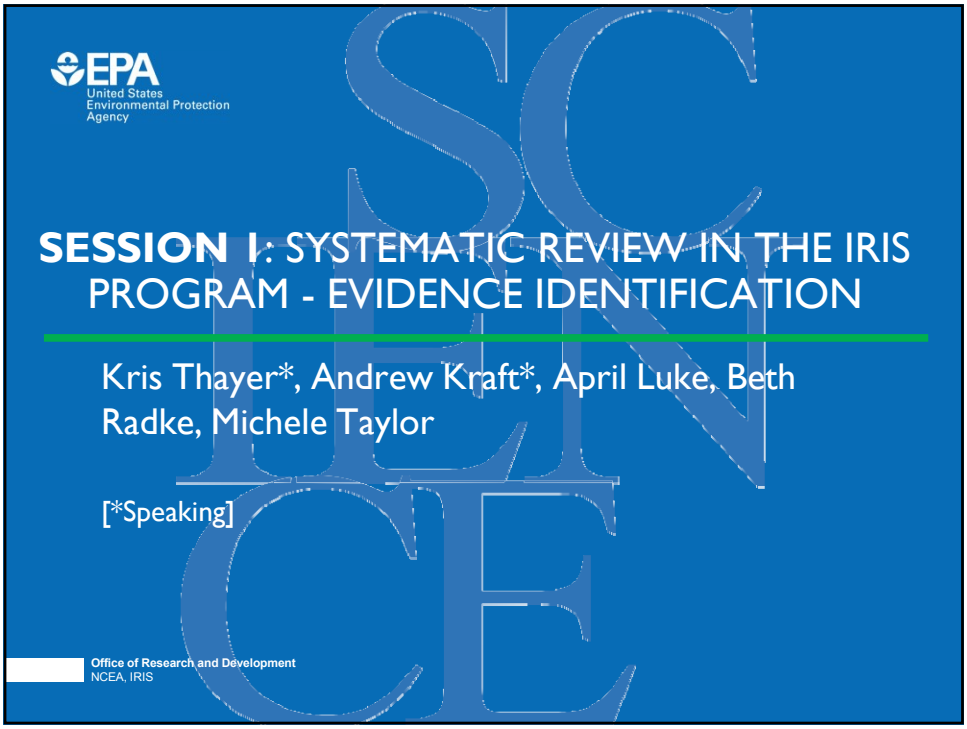
- **Increase transparency and full implementation of systematic review**
  - implement using approaches that foster consistency across the IRIS Program; many active and all new starts address systematic review-related recommendations of 2014 NAS report
- **Modernize the IRIS Program**
  - through automation and machine learning to expedite systematic review, incorporation of emerging data types
- **Modularize product lines**
  - implement a portfolio of chemical evaluation products that optimize the application of the best available science and technology. These products will allow IRIS to remain flexible and responsive to clients within the EPA as well the diverse collection of stakeholders beyond EPA, including states, tribal nations, and other federal agencies.
- **Enhance accessibility**
  - provide outreach and training to make systematic review practices ubiquitous and more accessible; enhance data sharing through publicly available software platforms for assessments developed by EPA, other federal and state agencies, industry, academia and other third-parties.

15

IRIS has Addressed the Major NAS 2014 Recommendations

NAS 2014 Topics	IRIS Process Improvements
General Process Issues (Chapter 2)	• Quality management pipeline implemented • Program and project management processes implemented • Frequent opportunities for stakeholder engagement
Future Directions (Chapter 8 "Lessons Learned" and "Looking Forward")	• Processes being implemented include flexibility to incorporate evolving methods in systematic review and risk assessment • Increased collaboration with federal partners and international experts prevents duplication of effort and maintains cutting edge approaches • Current research efforts and training serve to ensure that methods and staff are able to adapt to changing scientific contexts and sources of evidence, including new and emerging data types

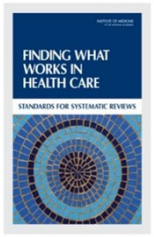
16



Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

Systematic Review

A structured and documented process for transparent literature review[1]



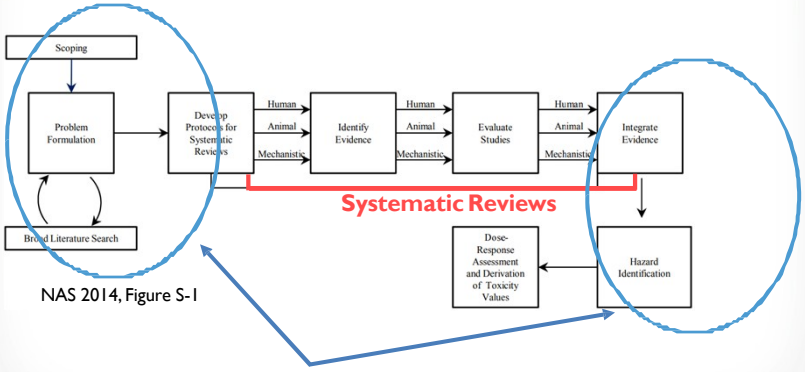
"As defined by IOM [Institute of Medicine], systematic review 'is a scientific investigation that focuses on a specific question and uses explicit, pre-specified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies.'" [p. 4] (NRC, 2014)

[1] Institute of Medicine. Finding What works in Health Care: Standards for Systematic Reviews. p.13-34. The National Academies Press. Washington, D.C. 2011

18

Systematic Review Elements (NAS 2014)

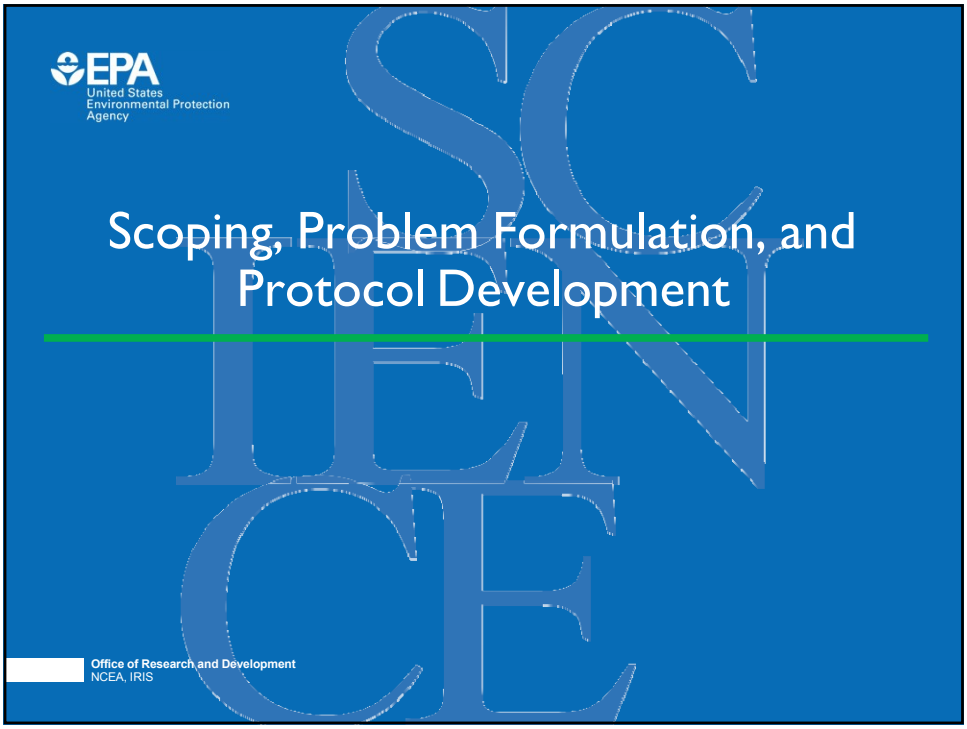
"In the context of IRIS, the committee has defined systematic review as including protocol development, evidence identification, evidence evaluation, and an analytic summary of the evidence"

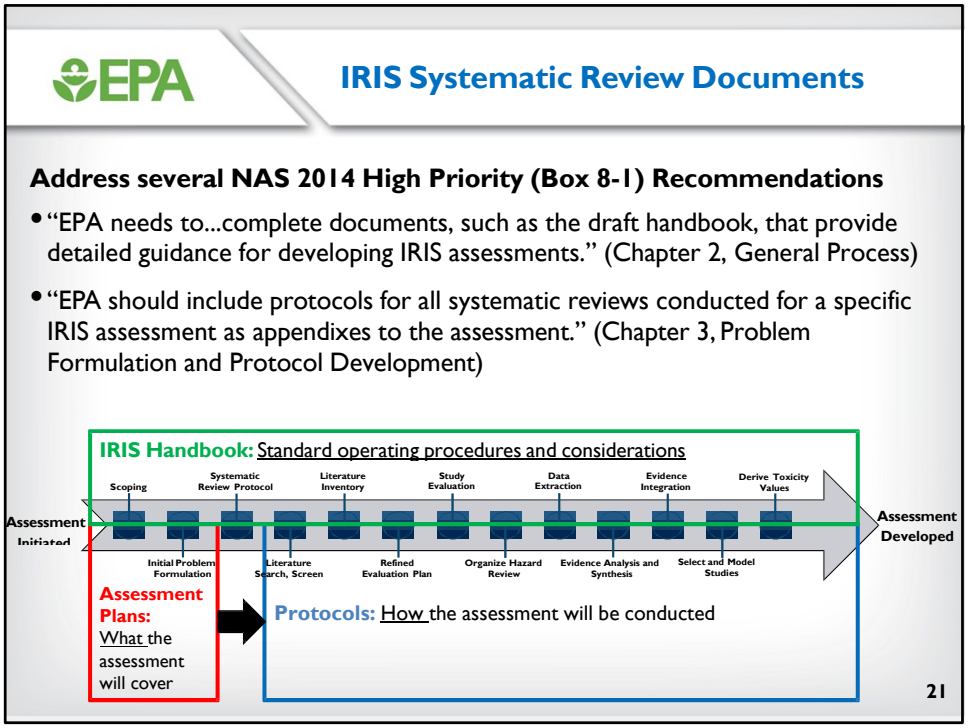


NAS 2014, Figure S-1

IRIS also considers these phases as part of its systematic review process

19





Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

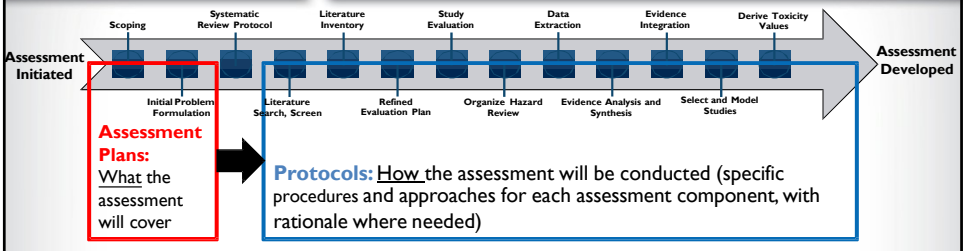
IRIS Handbook

IRIS Handbook: Approaches and considerations for applying principles of systematic review to IRIS assessments, general frameworks, and examples.

- IRIS Handbook level of detail aimed for EPA staff and contractors, e.g., use of HERO, timelines for internal review steps, etc.
- Currently being updated to reflect Agency input, evolving IRIS practices as systematic review approaches are tested through implementation, and public comment received on chemical-specific protocols (e.g., chloroform)
- Evergreen to reflect future advances
- Anticipate public release in 2018

22

IRIS Assessment Plans and Protocols



- Chemical-specific documents
- IRIS Assessment Plans (IAPs) are problem formulation and scoping documents that include more elements of systematic review
- Protocols outline methods, including updates to the IAPs
- IAPs and protocols include proposed "modularity," targeted focus and use of existing assessments
- Templates created to promote consistency across the IRIS Program, which is implemented across NCEA divisions and geographical locations

23

Appendix C

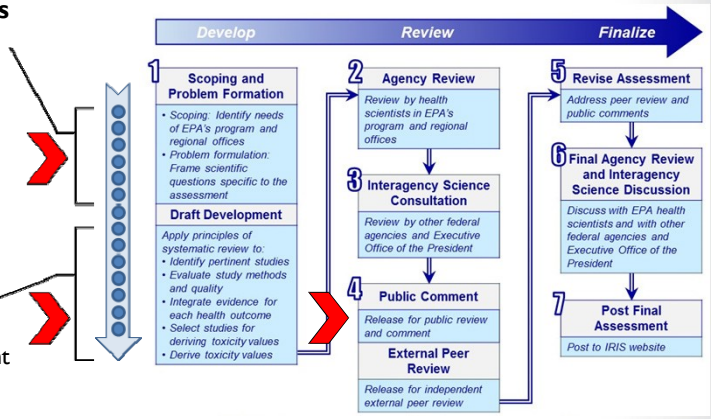
IRIS Assessment Plans, Protocols, and 7-Step IRIS Process

Early Step 1: IRIS Assessment Plans

- What the assessment covers
- 30-day public comment period + public science meeting

Mid-Step 1: Protocols

- How the assessment will be conducted
- 30-day public comment

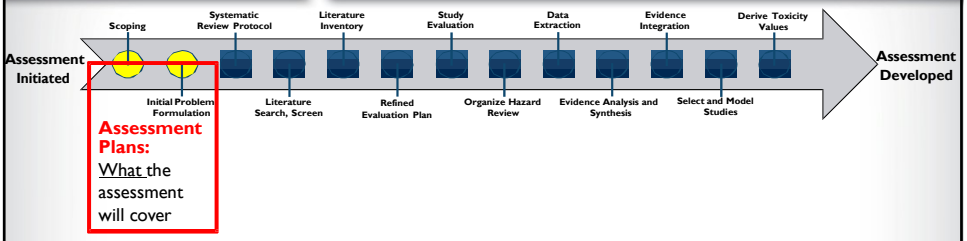


Opportunities for Public Comment

https://www.epa.gov/iris/basic-information-about-integrated-risk-information-system#process

24

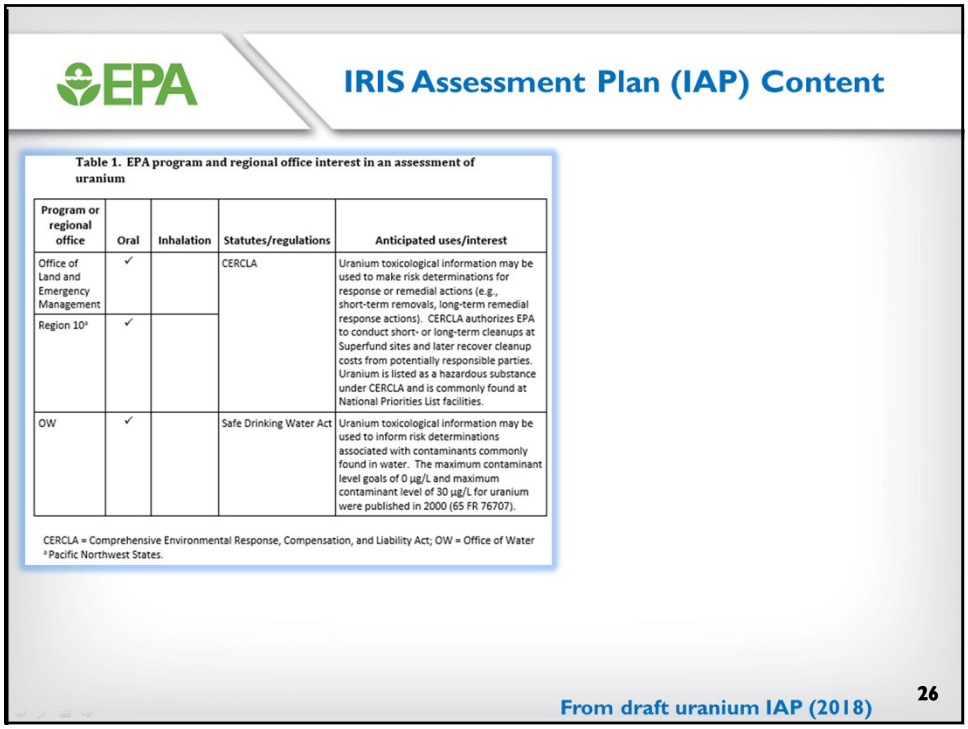
IRIS Assessment Plan (IAP)

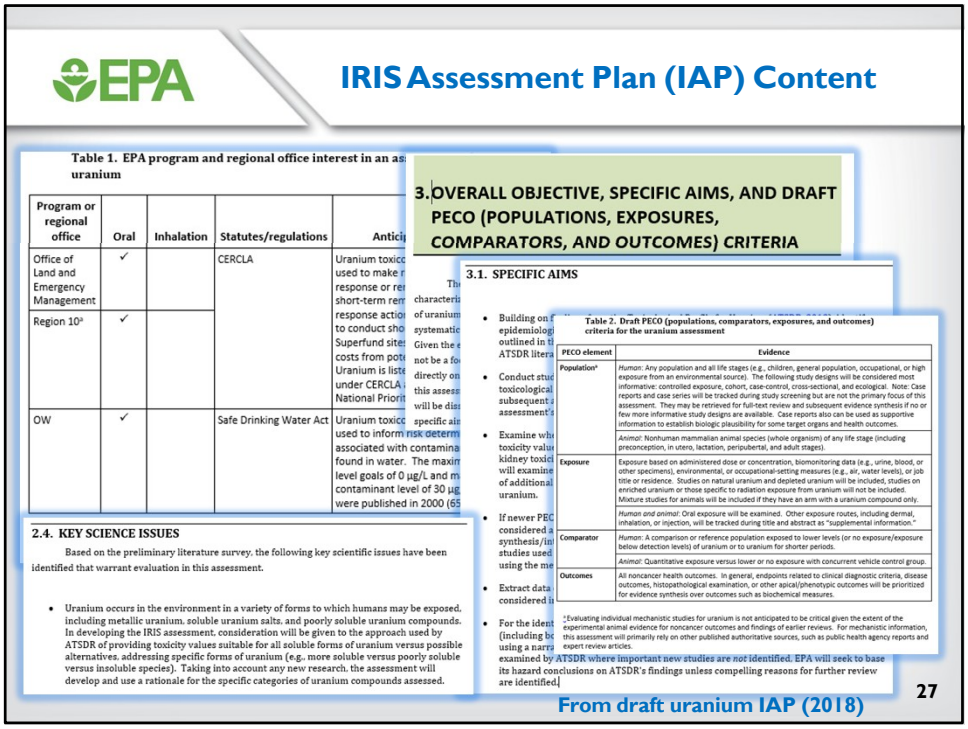


- Scoping and initial problem formulation determinations
  - Background and Agency need, exposure context, objectives and specific aims, key areas of scientific complexity
  - Includes draft PECO (Populations, Exposures, Comparators, and Outcomes) criteria which outlines evidence considered most pertinent
  - Internal review of IAP fosters early and focused Agency engagement
- Released for a 30-day public comment period + public science discussion (beginning of IRIS Step 1)
  - Examples: chloroform, ethylbenzene, nitrate/nitrite (Sept 2017), uranium (Jan 2018)

25

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation





IAP Can Include Literature Surveys

- Broad surveys to assess extent and nature of evidence, level of effort, type of expertise required
- Surveys inform decisions on targeted focus, e.g., evidence streams to consider core-PECO (versus supplemental), health outcomes likely covered in assessment
- Surveys may be developed based on other assessments, manual review of studies, or through use of specialized software applications

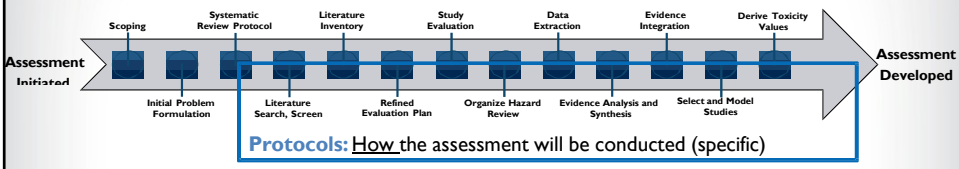
Nitrate/Nitrite (survey based on IARC 2010 and ATSDR 2017 assessments)

Outcomes	Human Studies				Animal Studies					
	Occupational epidemiology	General population epidemiology	Controlled exposure	Case reports and case series reports	Chronic	Subchronic	Short-term	Acute	Multi-generational	Gestational
Cancer		60			13					
Cardiovascular		1	1	3						
Dermal and ocular				1						
Developmental		14							2	6
Endocrine(thyroid)		6	1		4	3	1			
Gastrointestinal	1			7	5	1				
Hematological		25	3	10	4	6	3	1		
Hepatic					3			2		
Immunological										
Metabolic disease		8								
Musculoskeletal										
Neurological and sensory			1	6	1	1			1	
Renal					1					
Reproductive			3		2	2			1	
Respiratory										
Other					9	2	1		1	

The numbers represent the numbers of studies that investigated a particular health effect, and not the number of studies that identified a positive association with exposure.

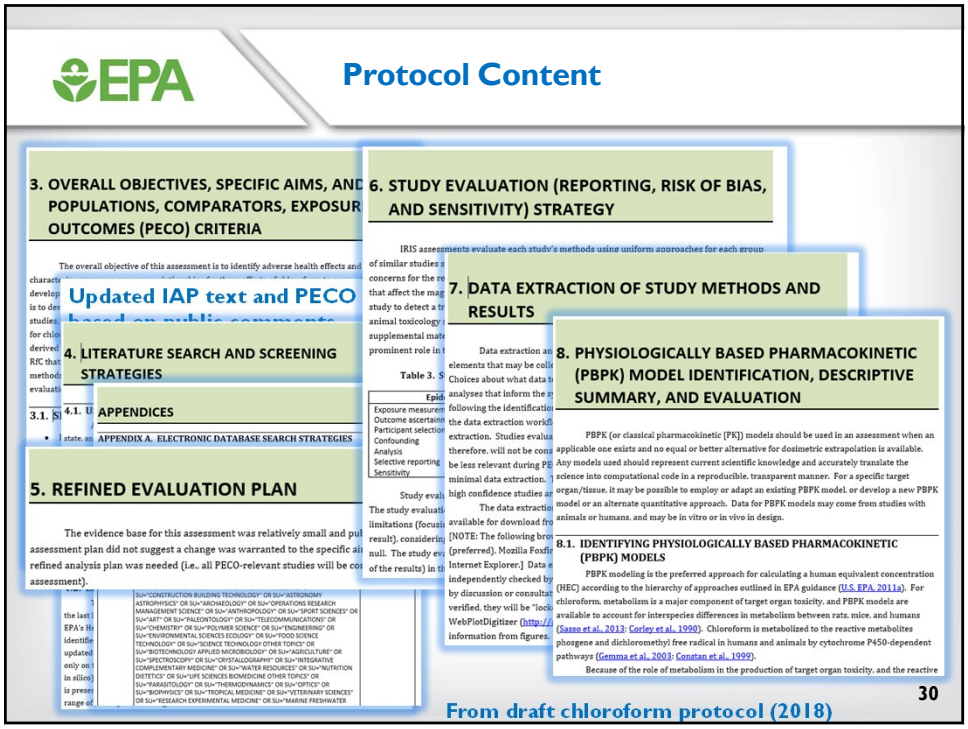
28

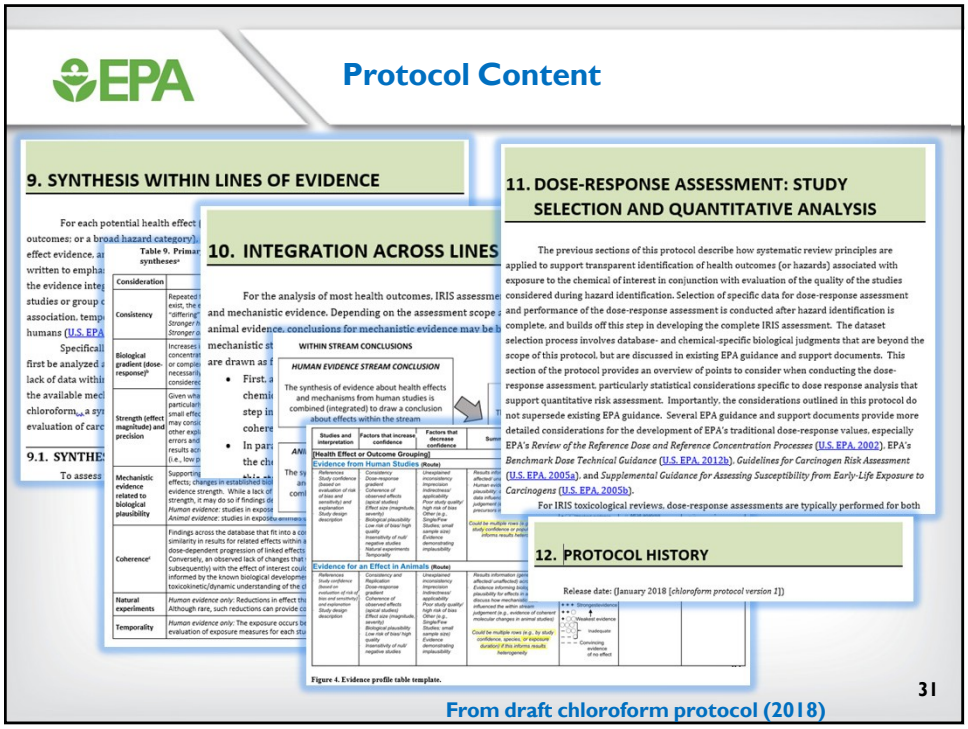
Protocols



- Assessment specific stand-alone method documents that do not rely on the IRIS Handbook to convey methodology
- Comments received on IAP are considered when preparing the protocol (updated IAP text is included in the protocol)
- Released for 30-day public comment period (during Step 1 of IRIS Process)
- List of included, excluded, and studies tagged as supplemental will be disseminated through protocols (either during initial release or as an update)
- Protocol is iterative - Knowledge gained during implementation may result in revisions to the protocol to focus on the best available evidence. Major revisions are documented via updates, e.g., changes to specific aims or PECO

29





EPA

Publicly Available Examples

Assessment Plans

September 27-28, 2017
- Chloroform
- Nitrate/nitrites
- Ethylbenzene

January 26, 2018
- Uranium

Protocol

January 26, 2018
- Chloroform (includes list of included studies)

Rapid systematic review
- EPA response to the Chloroprene Request for Correction (posted January 29, 2018)

- **Targeted focus:** chloroform, uranium, chloroprene
- **Modularity:** ethylbenzene
- **Use of existing assessments conducted by others:** nitrate/nitrate, uranium (ATSDR assessments)

- IAPs and/or protocols will be released for most in-progress assessments
  - Which document is released depends on extent of refinement in scope compared to previous public sharing and maturity of the draft assessment

32

EPA United States Environmental Protection Agency

Literature Searching, Screening, and Inventories*

Office of Research and Development
NCEA, IRIS

* includes basic methodological details

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

NAS 2014: High Priority (Box 8-1) Recommendations

"...include a section on evidence identification that is written in collaboration with information specialists trained in systematic reviews and that includes a search strategy for each systematic-review question being addressed in the assessment. Specifically, the protocols should provide a line-by-line **description of the search strategy, the date of the search, publication dates searched, and explicitly state the inclusion and exclusion criteria**..."

- **Protocols outline the specifics of the literature search and screening approaches, including inclusion and exclusion criteria in PECO tables**
- **Dedicated information technologists help formulate searches, and screening decisions are tracked in HERO (tagging)**
- **Manual and semi-automated approaches are being used to identify relevant studies**
- **Inventories of basic study methods organize evidence for refinement and evaluation**
- **Changes and updates are documented in the protocol**

34

Routine Evidence Identification Processes

Database Searches

- Identify peer-reviewed and "gray" (unpublished) literature
- PubMed, ToxLine, and Web of Science are standard (others can be included as needed)
- Conduct regular search updates
- Details of search strategy, dates, and retrieved records are presented in protocols and assessments

Screening

1. Title/abstract
2. Full text

- Use manual and automated approaches
- ≥ 2 screeners
- Tag studies as excluded, meeting PECO criteria, or supplemental information
- Screening decisions available in HERO
- Typically do not apply language-restrictions
- Review reference list of included studies and relevant reviews to identify studies missed from database searches
- Share list of included studies with public to further ensure all relevant studies included

Inventories

Health Outcome & PBPK Studies

- Tag studies by line of evidence and outcome
- Distribute to disciplinary experts for review

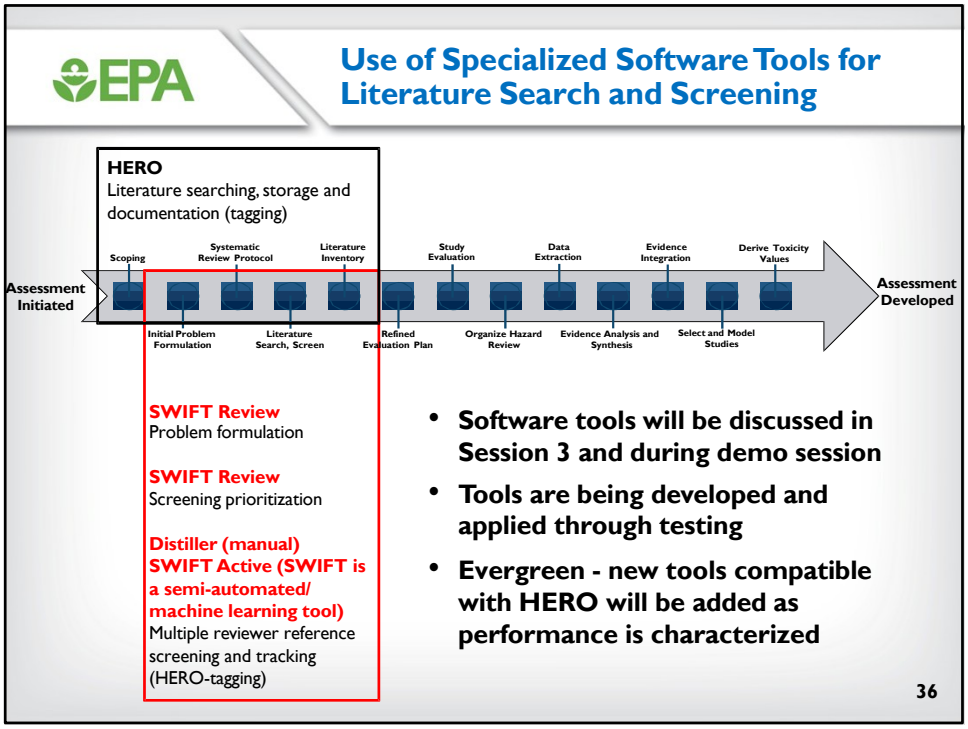
Supplemental Studies

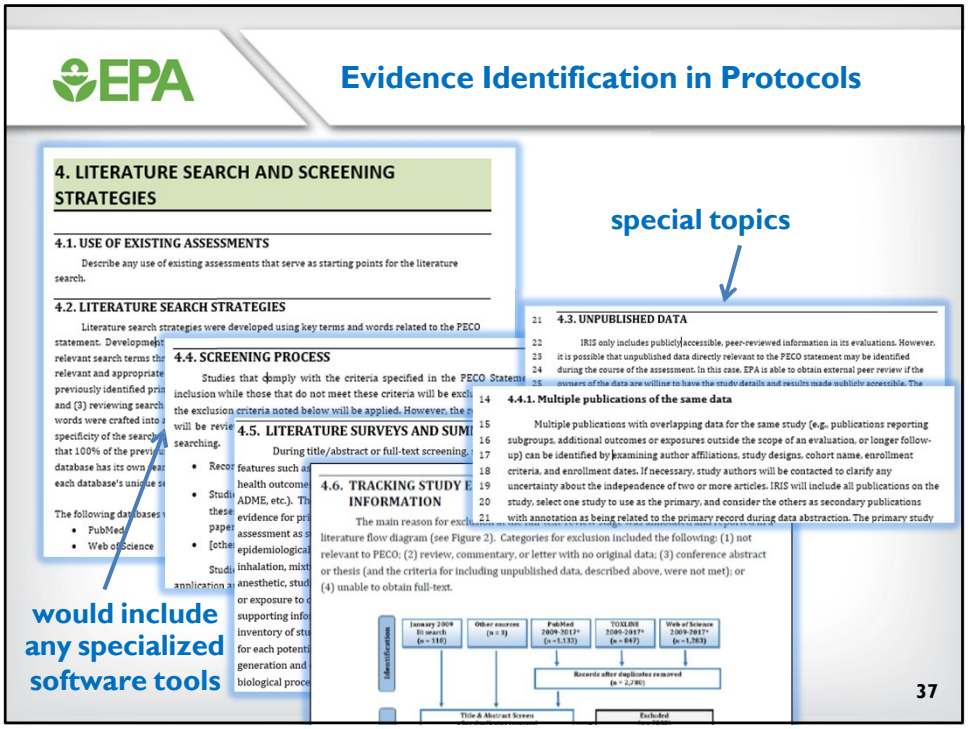
- Includes in vitro and other mechanistic evidence (e.g., non-PECO exposure route; non-PECO animal model; toxicokinetic data)

- Inventories contain basic study methods for evaluation and prioritization decisions

35

Appendix C





PECO Criteria to Identify Studies

PECO element	Evidence
Populations[a]	*Human*: Any population and lifestage (occupational or general population, including children and other sensitive populations). The following study designs will be considered most informative: controlled exposure, cohort, case-control, cross-sectional, and ecological. Note: Case reports and case series will be tracked during study screening, but are not the primary focus of this assessment. They may be retrieved for full-text review and subsequent evidence synthesis if no or few informative study designs are available. Case reports can also be used as supportive information to establish biologic plausibility for some target organs and health outcomes.
Exposures	Studies of chloroform in the context of its use as an anesthetic gas will be excluded.
	Animal: Any exposure to chloroform via inhalation. Studies employing chronic exposures or short-term, developmental-only exposures will be considered the most informative. Studies involving exposures to mixtures will be included only if they include an arm with exposure to chloroform alone. Studies utilizing chloroform as an extraction solvent to isolate specific chemical constituents will be excluded.
	Studies describing physiologically-based pharmacokinetic (PBPK) models for chloroform will be included.
Comparators	*Human*: A comparison or referent population exposed to lower levels (or no exposure/exposure below detection limits) of chloroform, or exposed to chloroform for shorter periods of time.
	Animal: A concurrent control group exposed to vehicle-only treatment.
Outcomes	All health outcomes (both cancer and noncancer). In general, endpoints related to clinical diagnostic criteria, disease outcomes, histopathological examination, or other apical/phenotypic outcomes will be prioritized for evidence synthesis over outcomes such as biochemical measures. As discussed above, based on preliminary screening work, EPA anticipates that a systematic review for health effect categories other than those identified (i.e., nasal cavity effects, nervous system effects, liver and kidney effects, immunotoxic effects, and reproductive/developmental effects) will not be undertaken unless a significant amount of new evidence is found upon review of references during the comprehensive literature search.

[a] Evidence from in vitro, in silico, and other types of mechanistic studies will be prioritized based on likelihood to impact evidence synthesis conclusions for human health. For chloroform, mechanistic studies will only be considered for evaluation if they have the potential of impacting the existing 2001 MOA analysis, or are essential for answering questions identified during the human and animal evidence syntheses.

Example from the draft chloroform protocol

38

Example Literature Screening Form

SUBMIT FORM and go to ▾ or Skip to Next

Forms Independently Entered by 2 Reviewers

1. Based on Title and Abstract does the article contain relevant human, animal, or in vitro evidence?
   ○ Yes ○ No ◉ No, but has supportive information
   ○ Unclear (e.g., no abstract) Clear Response

2. What kind of evidence or supportive information?
   ☐ human
   ☐ animal
   ☑ in vitro, omics, alternative model systems

3. What kind of supportive information?
   ☑ MOA/mech (cancer)
   ☐ MOA/mech (non-cancer)
   ☐ case report or poisoning
   ☐ non-inhalation route
   ☐ mixture
   ☐ ADME/PBPK
   ☐ exposure assessment
   ☐ THM, disinfection/chlorination by-product, swimming pools
   ☐ susceptible population
   ☐ anesthesia/inhalent

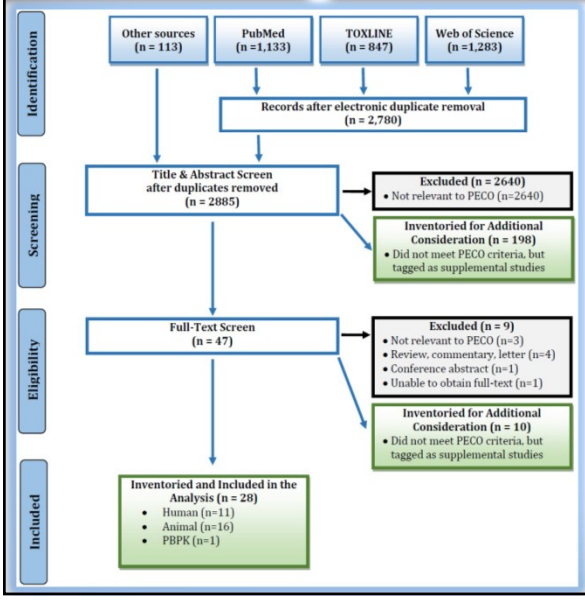
P	Human: Any population and life stage (occupational or general population, including children and other sensitive population). The following study designs will be considered most informative: controlled exposure, cohort, case-control, cross-sectional, and ecological. Note: Case reports and case series will be tracked during study screening, but are not the primary focus of this assessment. They may be retrieved for full-text review and subsequent evidence synthesis if no or few informative study designs are available. Case reports can also be used as supportive information to establish biologic plausibility for some target organs and health outcomes. Animal: Nonhuman mammalian animal species (whole organism) of any life stage (including preconception, in utero, lactation, peripubertal, and adult stages).
E	Human: Any exposure to chloroform, including occupational exposures, via inhalation. Exposures quantified by either actual exposure measurements or occupational exposure history are preferred. Studies of chloroform in the context of its use as an anesthetic gas will be excluded. Animal: Any exposure to chloroform via inhalation. Studies employing chronic exposures or short-term, developmental-only exposures will be considered the most informative. Studies involving exposures to mixtures will be included only if they include an arm with exposure to chloroform alone. Studies utilizing chloroform as an extraction solvent to isolate specific chemical constituents will be excluded. PBPK: Studies describing physiologically-based pharmacokinetic (PBPK) models for chloroform will be included.
C	Human: A comparison or reference population exposed to lower levels (or no exposure/exposure below detection limits) of chloroform, or exposed to chloroform for shorter periods of time. Animal: A concurrent control group exposed to vehicle-only treatment.
O	All health outcomes (both cancer and noncancer). In general, endpoints related to clinical diagnostic criteria, disease outcomes, histopathological examination, or other apical/phenotypic outcomes will be prioritized for evidence synthesis over outcomes such as biochemical measures. As discussed above, based on preliminary screening work, EPA anticipates that a systematic review for health effect categories other than those identified (i.e., nasal cavity effects, nervous system effects, liver and kidney effects, immunotoxic effects, and reproductive/developmental effects) will not be undertaken unless a significant amount of new evidence is found upon review of references during the comprehensive literature search.

SUBMIT FORM and go to ▾ Skip to Next

Draft example based on chloroform using Distiller

39

Tracking: Literature Flow Diagrams



- **Track rationale for full-text exclusions**
- **Use HERO to share repositories of included, excluded, and supplemental studies**

Example modeled on the draft chloroform protocol

40

Literature Inventories

Example Details Routinely Extracted (*female reproductive toxicity in animals*):

- Outcome category (e.g., fertility) and/or Specific endpoint (e.g., number of litters)
- Species (e.g., rat; alternative [nonmammalian] animal)
- Exposure duration (e.g., chronic; multi-generational; gestational)
- Exposure route (e.g., oral [gavage]; in vitro)

Assessment-Specific Extraction Details (*generic examples*):

- Exposure levels tested
- Test article details, such as purity or isomeric composition

Results are Typically <u>Not</u> Included in Inventories

Developing Extraction Forms *(all 3 lines of evidence) to be interoperable with HAWC*

Test compound	DBP	Maternal age at init. gestational exp.	NA
Species	Rat	Subject age at in vivo exposure initiation	PND21
Exposure type	In vivo	Lifestage at in vivo exposure initiation	Weaning
In vivo exposure route	Oral-gavage	Dose/concentration	0, 500 mg/kg-day
Strain of exposed test model	Sprague-Dawley	Exposure duration	Single exposure
Sex of exposed test model	M	Exposure period	Postnatal only

41

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

Refined Evaluation Plan (optional)

Discipline-specific experts consider whether and how to further refine or prioritize studies/outcomes for evaluation (based on study design features)

- ***Health effect studies meeting PECO criteria (e.g., organized by outcome):***
  - Considers ADME and other key science issues (supplemental studies reviewed)
  - Opportunity to discuss outcome grouping (e.g., based on known biology/MOA) and handling of key science issues during outcome-specific study evaluations
  - Studies with certain design features or specific outcomes may be selected or prioritized for evaluation and synthesis (e.g., based on exposure duration, administration, or levels tested; or endpoint specificity)
- ***Supplemental mechanistic studies (e.g., organized by test system, mechanistic event, or key characteristic [of carcinogens]) are considered iteratively:***
  - Identifies other studies on specific aim mechanistic questions (e.g., mutagenicity)
  - Organizes the available evidence to allow for pragmatic evaluations of key issues that arise during review of PECO-specific human and animal studies (Session 2)

Refinements are tracked and updated in the assessment protocol 42

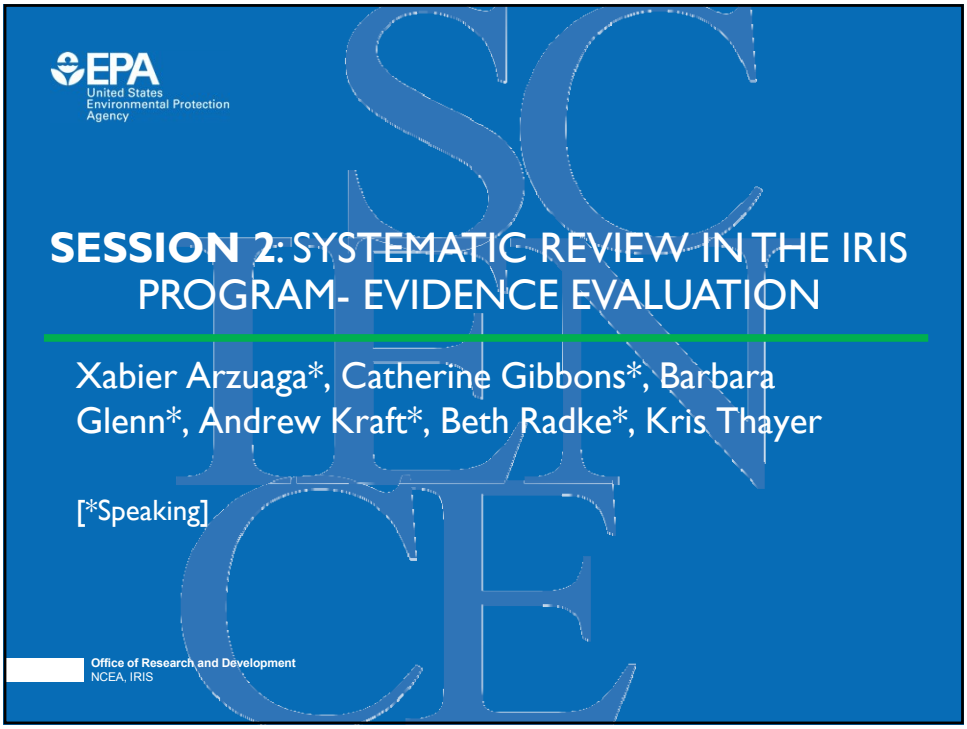
IRIS has Addressed the Major NAS 2014 Recommendations

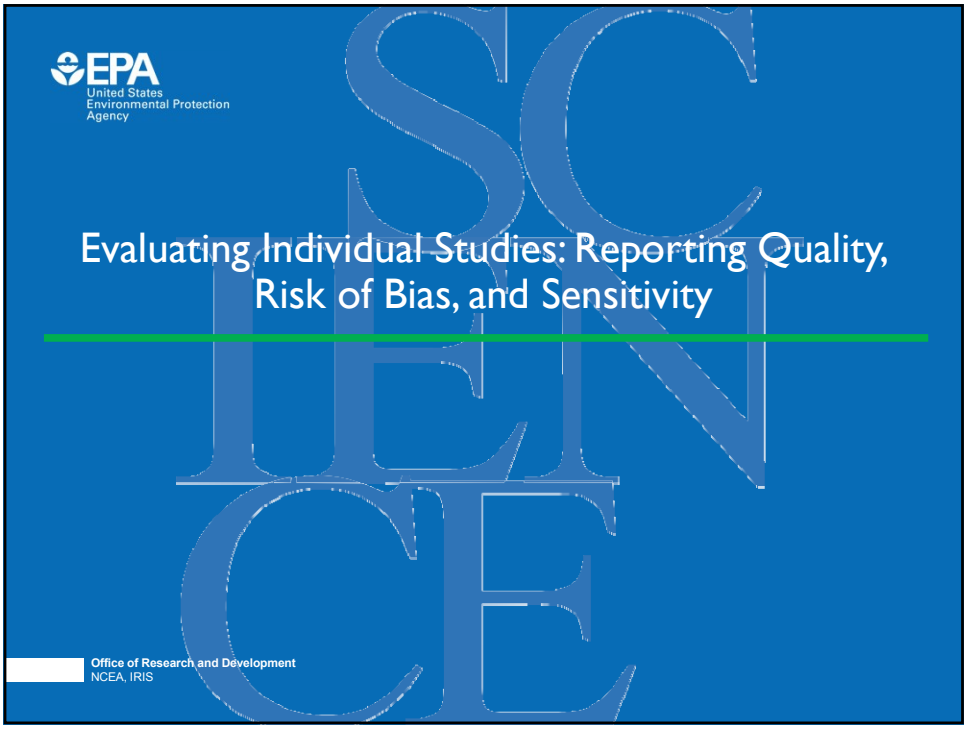
NAS 2014 Topics	IRIS Process Improvements
General Process Issues (Chapter 2); Problem Formulation and Protocol Development (Chapter 3)	• Draft IRIS Handbook of program SOPs is being reviewed within EPA • IAPs allow early comment on problem formulation • More frequent Agency engagement facilitates scope refinement • Assessment protocols describe methods and allow for iteration • Re-occurring staff training and template IAPs and protocols promote consistency and quality control
Evidence Identification (Chapter 4)	• Consultation with information technologists and subject experts • Adopts current systematic review best practices, including use of specialized tools • Transparent documentation (e.g., literature flow diagrams)

See Demonstrations:

- Sciome Workbench for Interactive computer-Facilitated Text mining (SWIFT Review and SWIFT Active)
- Health Assessment Workspace Collaborative (HAWC)
- Heath Effects Research Online (HERO)

43





Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

NAS 2014 High Priority (Box 8-1) Recommendations on Evidence Evaluation

"When considering any method for evaluating individual studies, EPA should **select a method that is transparent, reproducible, and scientifically defensible**. Whenever possible, there should be empirical evidence that the methodologic characteristics that are being assessed in the IRIS protocol have systematic effects on the direction or magnitude of the outcome."

"EPA should **specify the empirically based criteria it will use to assess risk of bias for each type of study design in each type of data stream.**"

"To maintain transparency, EPA should **publish its risk-of-bias assessments as part of its IRIS assessments**."

46

Study Evaluation – Developing an Approach

- **Considered and drew from existing tools for study evaluation.**
- **Developed approaches for both epidemiology and toxicity studies that:**
  - **Addresses study sensitivity and identifies potential sources of bias.**
  - **Transparently presents the criteria/considerations used to consistently evaluate and judge each study/outcome.**
  - **Provides access to the rationale for discipline-specific decisions made during the evaluation process.**
- **Objective of the approach: Identify the most informative and reliable studies for evidence synthesis and integration.**

47

Appendix C

PBPK Model Evaluation

Criteria	Example information
Scientific	Biological basis for the model is accurate • e.g., Predicts dose metrics expected to be relevant
	Consideration of model fidelity to the biological system strengthens the scientific basis relative to standard extrapolation (default) approaches • e.g., Can the model describe critical behavior, such as nonlinear kinetics in a relevant dose range, better than the default (i.e., $BW^{3/4}$ scaling)?
	Principle of parsimony (i.e., model complexity or biological scale should be commensurate with data available to identify parameters)
	Model describes existing PK data reasonably well, both in "shape" (e.g., matches curvature) and quantitatively (e.g., within a factor of 2−3)
	Model equations are consistent with biochemical and biological understanding
Initial technical	Well-documented model code is readily available to EPA and public
	Set of published parameters clearly identified, including origin/derivation
	Parameters do not vary unpredictably with dose • e.g., Any dose dependence in absorption constants is predictable across the dose ranges relevant for animal and human modeling
	Sensitivity and uncertainty analysis has been conducted for relevant exposure levels (local sensitivity analysis is sufficient, though global preferred) • e.g., A sound explanation should be provided when sensitivity of the dose metric to model parameters differs from what is reasonably expected

Prior to use, relevant PBPK models will:

- Be thoroughly evaluated based on scientific and technical criteria (examples to the left).
- Undergo QA/QC on model equations, parameters (including primary/secondary sources), and model code.

For details, please see:

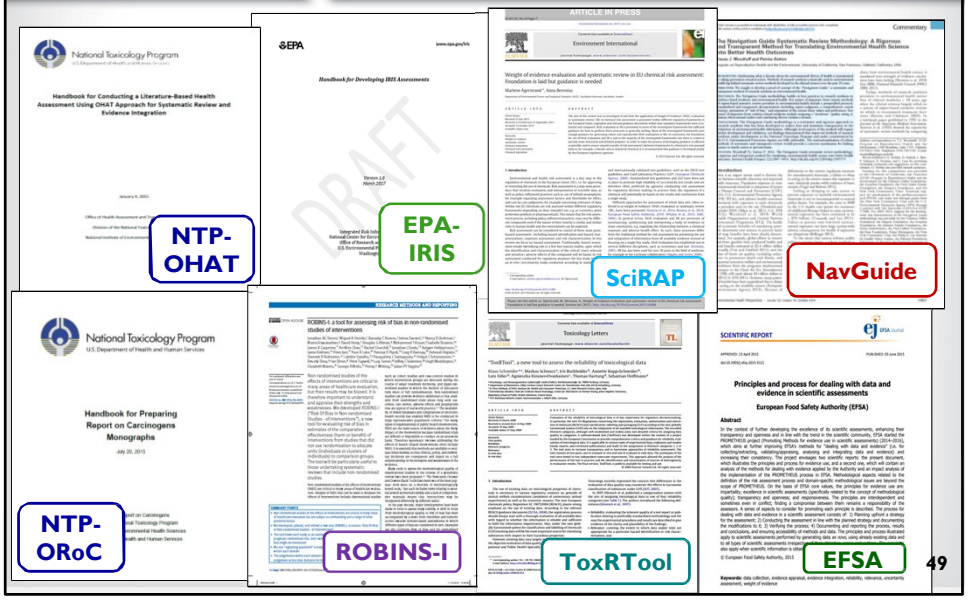
- Poster:

  Systematic evaluations of PBPK models for human health risk assessment
- EPA website:

  EPA Response to the Request for Correction of the IRIS Toxicological Review of Chloroprene (2018)

48

Evolving Approaches



49

Evaluation of Individual Health Effect Studies

- General approach same for human and animal studies
- Evaluation process focused on:
  - Internal validity/bias
  - Sensitivity
  - Reporting quality

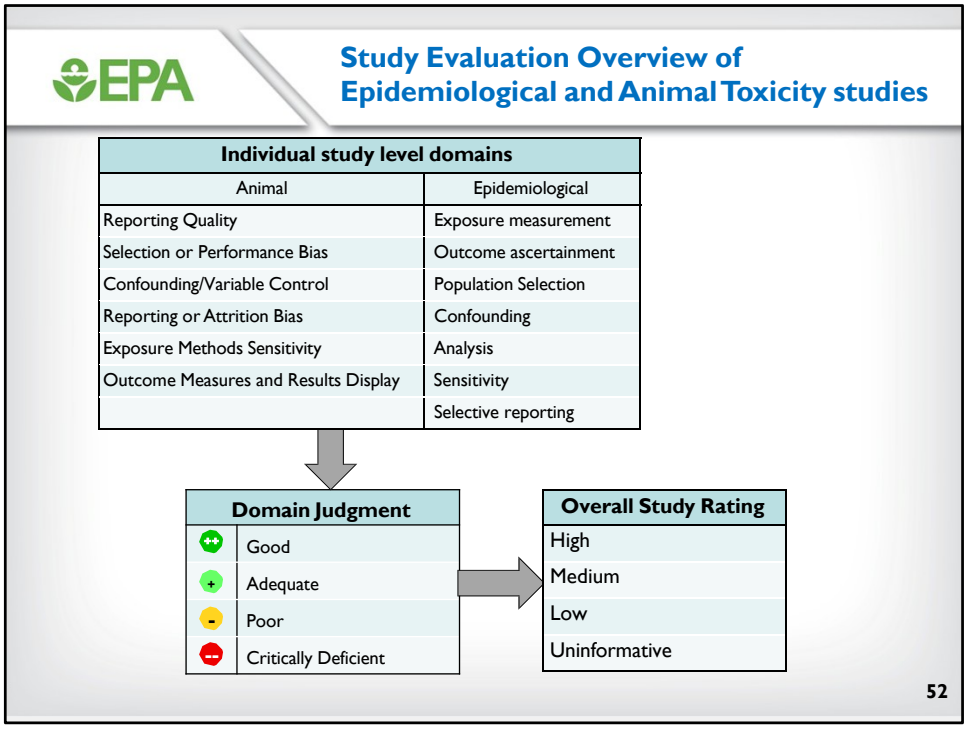
50

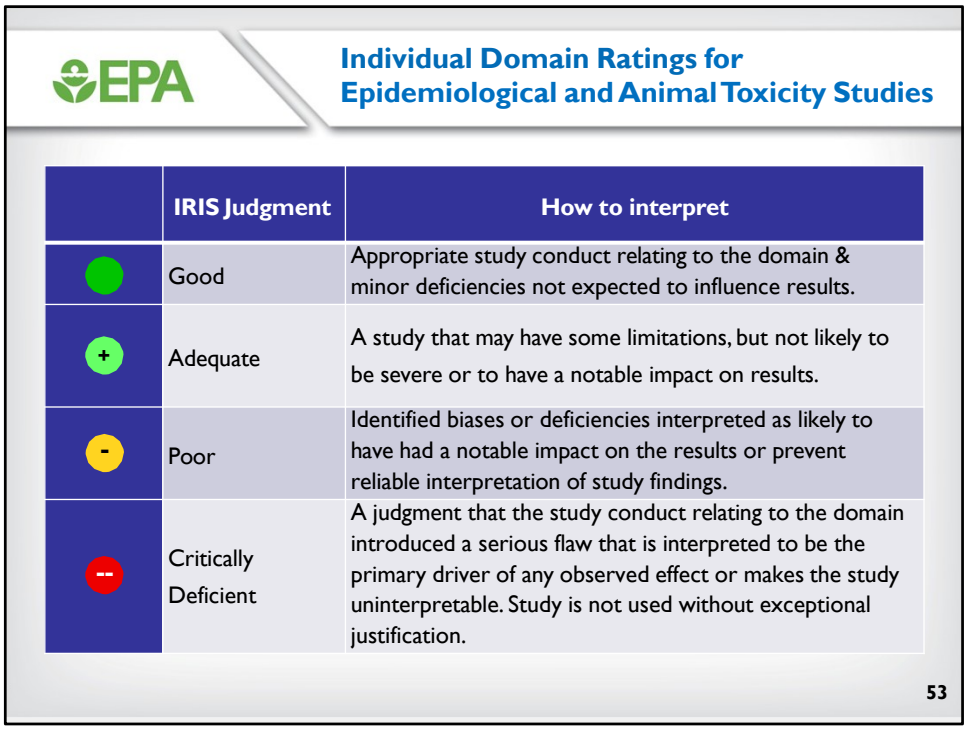
Development of Evaluation Strategies

- Questions in IRIS Protocol Template highlight general study attributes or elements to consider
- Subject-matter knowledge is used to formulate a list of issues to consider in the evaluation
- Develop a set of considerations based on exposure and outcome-specific knowledge

51

Appendix C





Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

Overall Study Confidence Ratings for Epidemiological and Animal Toxicity Studies

Rating	Description
High	No notable deficiencies or concerns identified; potential for bias unlikely or minimal and sensitive methodology.
Medium	Possible deficiencies or concerns noted, but resulting bias or lack of sensitivity would be unlikely to be of a notable degree.
Low	Deficiencies or concerns were noted, and the potential for substantive bias or inadequate sensitivity could have a significant impact on the study results or their interpretation.
Uninformative	Serious flaw(s) makes study results unusable

54

General Considerations to Evaluate Outcomes from Animal Toxicology Studies

Domain	Metric
Reporting Quality	Reporting of information necessary for study evaluation
Selection or Performance Bias	Allocation of animals to experimental groups
	Blinding of investigators, particularly during outcome assessment
Confounding/Variable Control	Control for variables across experimental groups
Reporting or Attrition Bias	Lack of selective data reporting and unaccounted for loss of animals
Exposure Methods Sensitivity	Characterization of the exposure to the compound of interest
	Utility of the exposure design for the endpoint of interest
Outcome Measures and Results Display	Sensitivity and specificity of the endpoint evaluations
	Usability and transparency of the presented data

55

Epidemiology Study Evaluation

- **Approach based on the Cochrane Risk of Bias in Nonrandomized Studies of Interventions (ROBINS-I)[1], modified for environmental and occupational exposures**
- **Start by considering an "ideal" study for each domain, identifying "critical deficiencies", then developing criteria to define other levels of confidence**
- **Emphasis is on discerning bias that would produce a substantive change in the estimated effect estimate.**

[1]Sterne, Hernan, et al. ROBINS-I: a tool for assessing risk of bias in non-randomized studies of interventions. BMJ 2016; 355:i4919.

56

Epidemiology Evaluation Domains

Domain	Core Question
Exposure measurement	Does the exposure measure reliably distinguish between levels of exposure in an appropriate time window?
Outcome ascertainment	Does the outcome measure reliably distinguish the presence or absence (or degree of severity) of the outcome?
Population selection	Is there evidence that selection into or out of the study (or analysis sample) was jointly related to exposure and outcome?
Confounding	Is confounding of the effect of the exposure likely?
Analysis	Does the analysis strategy and presentation convey the necessary familiarity with the data and assumptions?
Sensitivity	Are there concerns for study sensitivity?

57

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

Example of Considerations by Domains

Domain	Core Question
Exposure measurement	Does the exposure measure reliably distinguish between levels of exposure in an appropriate time window?

Examples of Prompting Questions:

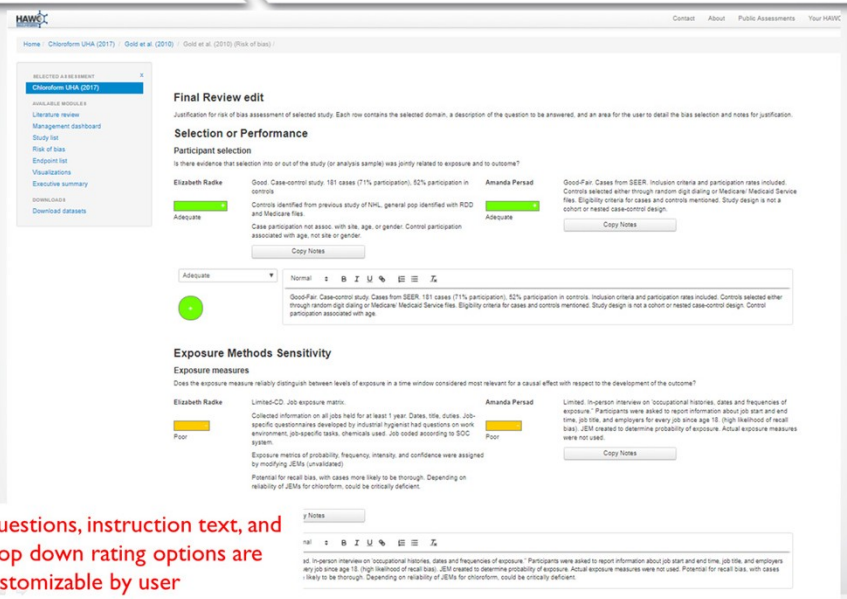
- Does the exposure measure capture the variability in exposure among the participants, considering intensity, frequency, and duration of exposure?
- Does the exposure measure reflect a relevant time window?
- Was exposure measurement likely to be affected by knowledge of outcome or by presence of the outcome (i.e., reverse causality)?

Examples of Follow-up Questions:

- Is the degree of exposure misclassification likely to vary by exposure level?
- If there is a concern about the potential for bias, what is the predicted direction of the bias on the effect estimate?

58

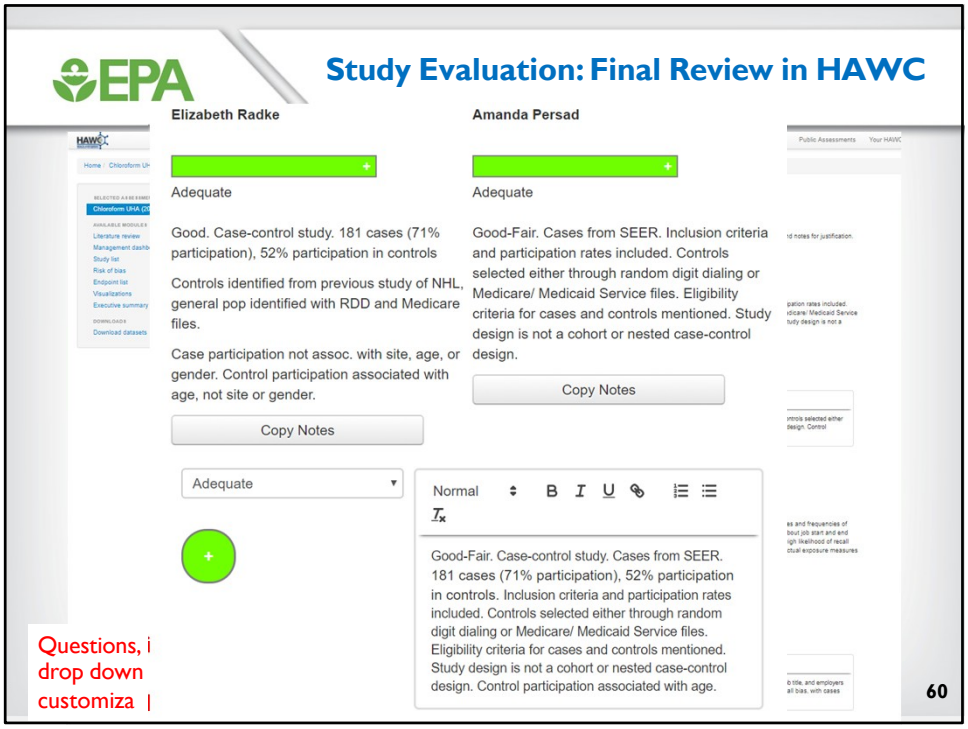
Study Evaluation: Final Review in HAWC

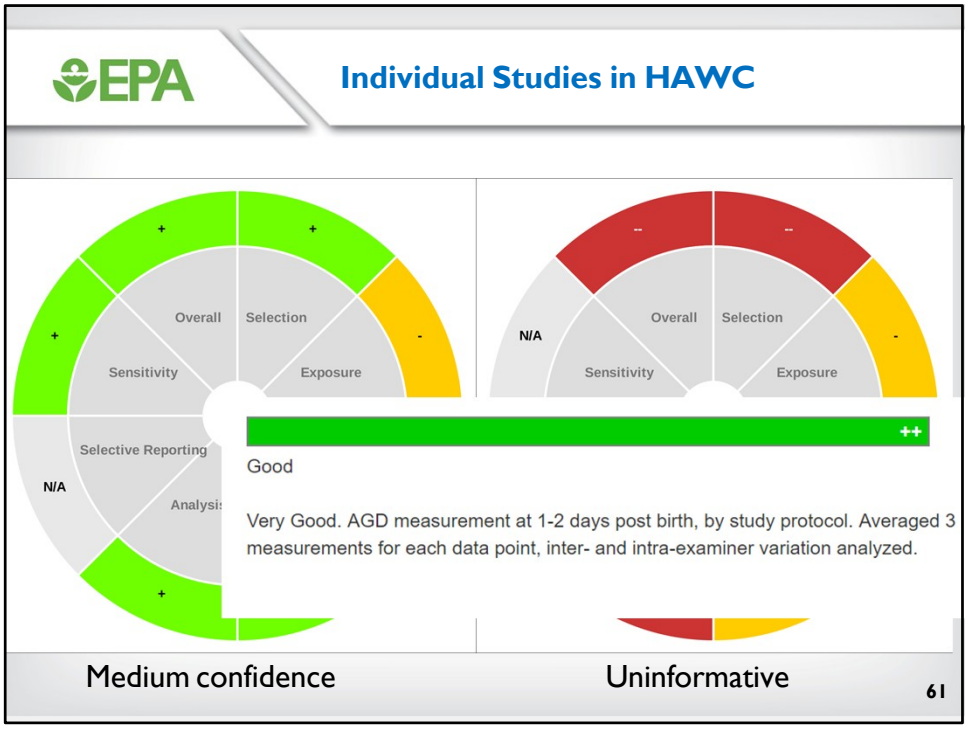


Questions, instruction text, and drop down rating options are customizable by user

58

Appendix C





Study Evaluation Summary in HAWC

	Study 1	Study 2	Study 3	Study 4	Study 5	Study 6
Population selection	++	+	++	-	++	+
Exposure measurement	-	-	-	-	-	-
Outcome ascertainment	++	++	++	+	-	++
Confounding	+	+	+	-	+	+
Analysis	++	++	++	+	-	+
Sensitivity	+	-	+	-	-	+
Overall study confidence	M	L	M	L	L	M

Legend

Domain judgement		Overall study rating	
++	Good	H	High confidence
+	Adequate	M	Medium confidence
NR	Not reported		
-	Poor	L	Low confidence
--	Critically deficient	U	Uninformative
N/A	Not applicable		

62

Publicly available examples

- **Initial and iterative improvements to study evaluation**
  - **Ammonia, Inhalation (final 2016)**
  - **RDX (peer review draft 2016)**
  - **TBA (peer review draft 2017)**
  - **ETBE (peer review draft 2017)**
- **Current methods for study evaluation**
  - **Chloroform protocol (2018)**
  - **EPA Response to Chloroprene Request for Correction (2018)**

63

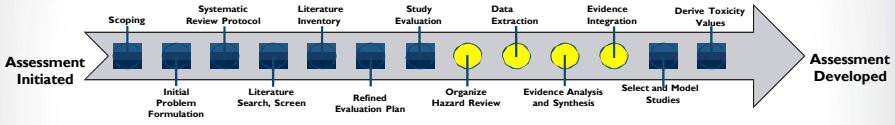
EPA United States Environmental Protection Agency

Evaluating Confidence in a Body of Evidence: Evidence Synthesis and Integration to Reach Hazard Conclusions

Office of Research and Development
NCEA, IRIS

EPA

Synthesis and Integration of Evidence Linking Exposure and Health Effects: Purpose



Synthesis: To describe the types of information within each line of evidence (human, animal and mechanistic), and to analyze and present study results regarding a given health effect to facilitate integration judgments.

- Decisions about the organization of the synthesis made prior to data extraction
- Narratives, but not study by study summaries
- Highlight information that informs the hazard evaluation

Integration: To develop judgments regarding strength of evidence for a health effect across lines of evidence

- A two-step process involving transparent and structured approaches for drawing summary conclusions across lines of evidence

65

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

NAS 2014: Relevant Comments and Recommendations

The NAS 2014 report discusses the complexities with organizing analyses around mechanism, noting that, "The history of science is replete with solid causal conclusions in advance of solid mechanistic understanding." (NRC, 2014, p. 90).

- The current approach focuses first on the available human and animal studies on health effects, incorporating mechanistic information at various stages of assessment development to clarify identified gaps in understanding (e.g., human relevance of animal-model data).

"The risk-of-bias assessment of individual studies should be carried forward and incorporated into the evaluation of evidence among data streams." (NAS 2014 Recommendation, Box 8-1)

- The results of the evaluation of individual studies is a critical component of the current evidence synthesis processes and integration frameworks.

66

NAS 2014: Relevant High Priority (Box 8-1) Recommendations

"**EPA should continue to improve its evidence-integration process incrementally and enhance the transparency of its process**. It should either maintain its current guided-expert-judgment process but make its application more transparent or adopt a structured (or GRADE-like) process...the committee does not offer a preference but suggests that EPA consider which approach best fits..."

"EPA should expand its ability to perform quantitative modeling of evidence integration."

- The current approach continues to use a guided expert judgment process, but structured sets of categorical criteria for decision-making within that process are more explicitly defined.
- The current frameworks, and documentation of decisions within these frameworks, enhance transparency, reproducibility, and comparability across health effects and assessments; these approaches are evolving within NCEA and across the field.
- Current research activities include quantitative methods to integrate evidence across streams (e.g., Bayesian approaches; see Session 4)

67

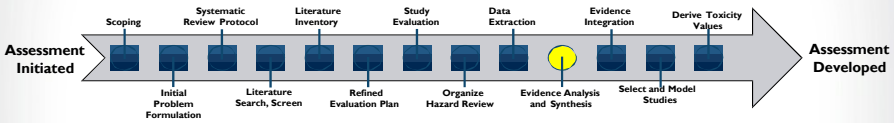
Synthesizing Evidence on Health Effects – Organization and Structure

Some questions about the evidence

- What outcomes are relevant to each health hazard domain and at what level (e.g., health effect or subgroupings) should synthesis occur?
- What populations were studied (e.g., general population, occupations, life stages, species, etc.) and do responses vary?
- Can study results be described across varying exposure patterns, levels, duration or intensity?
- Are there differences in the confidence in study results for different outcomes, populations, or exposure?
- Does toxicokinetic information explain differences in responses across route of exposure, other aspects of exposure, species, or life stages?
- How might dose response relationships be presented (specific study results or across study results)?

68

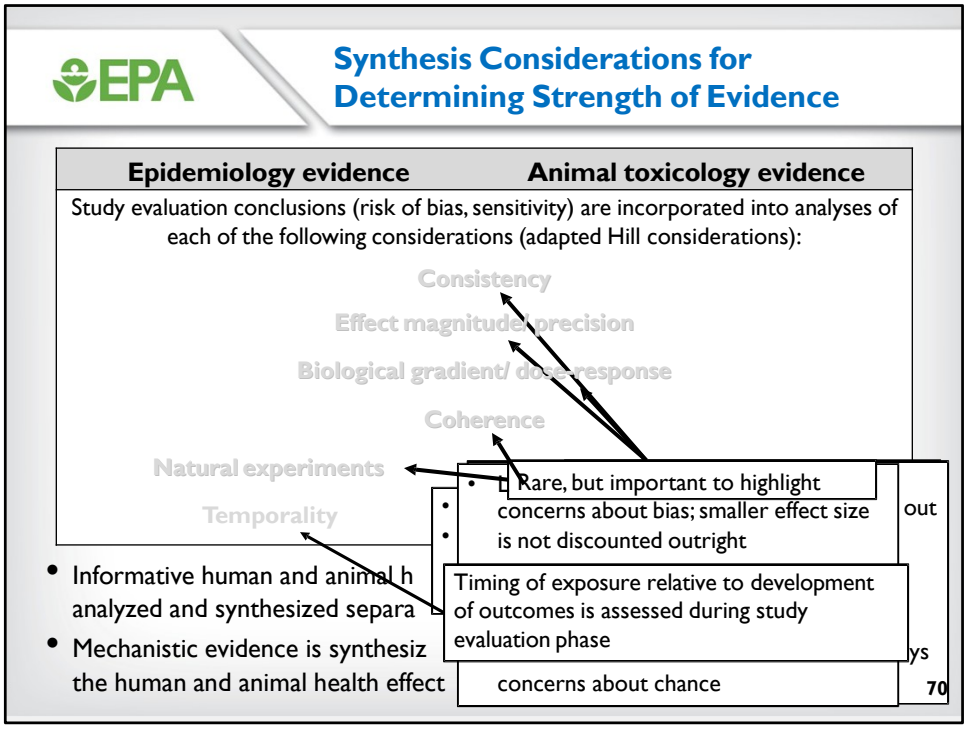
Scientific Judgment in Analysis and Synthesis of Evidence

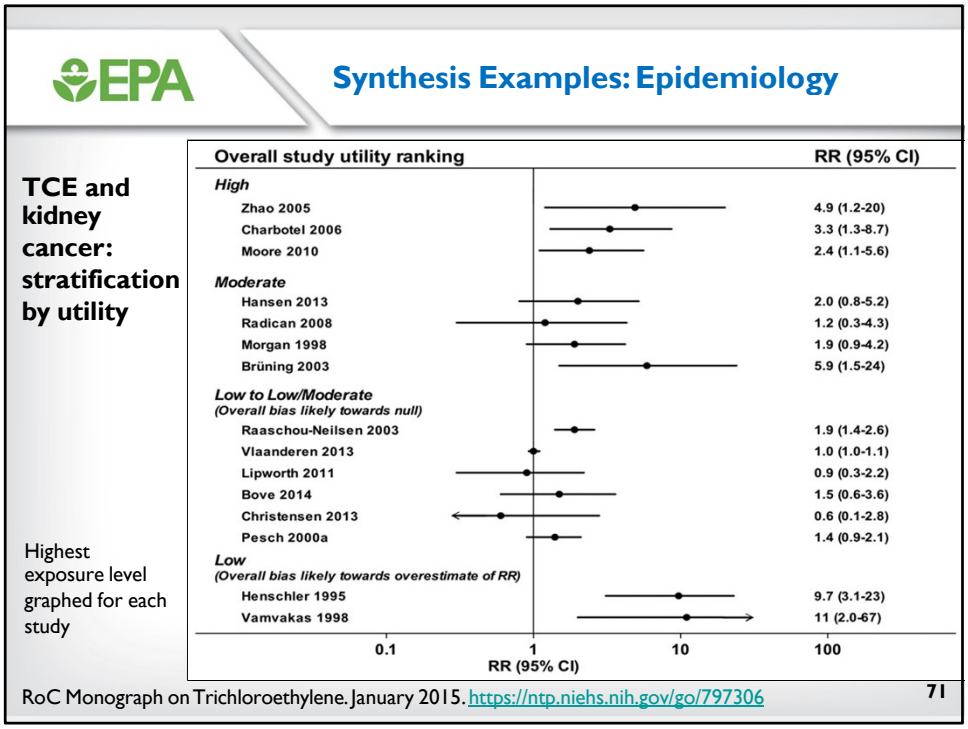


- Synthesis of evidence is more than counting the number of "positive" and "negative" studies
- Must systematically consider the influence of bias and sensitivity when describing study results and synthesizing evidence
- Synthesis should primarily be based on studies of medium and high confidence (when available)
- Analysis should try to draw conclusions about the strength of evidence from findings across collections of studies

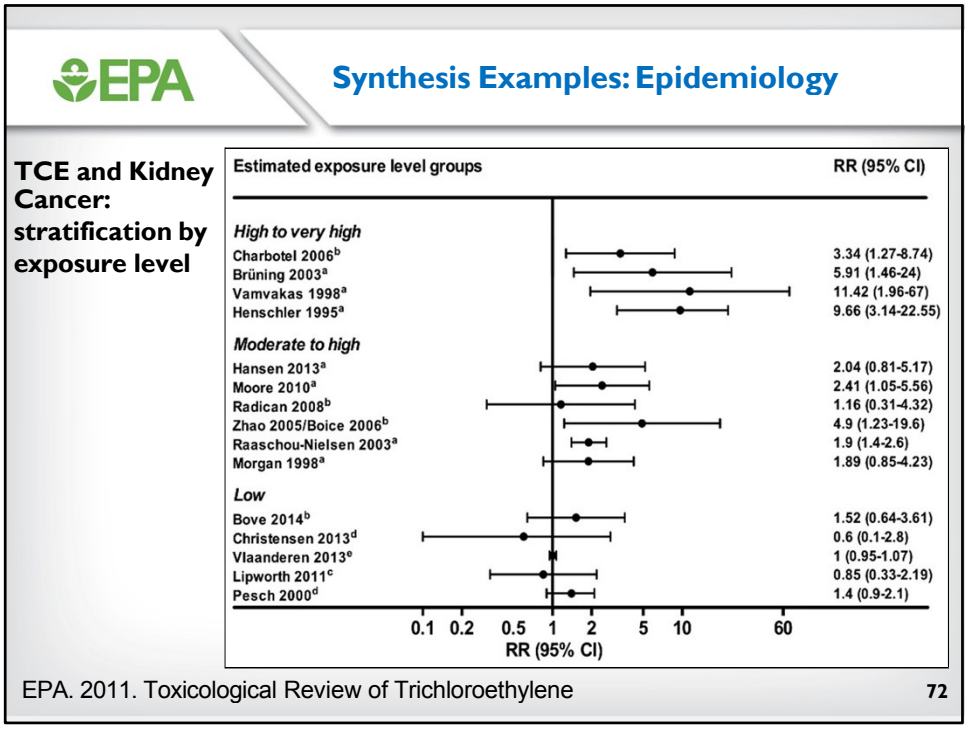
69

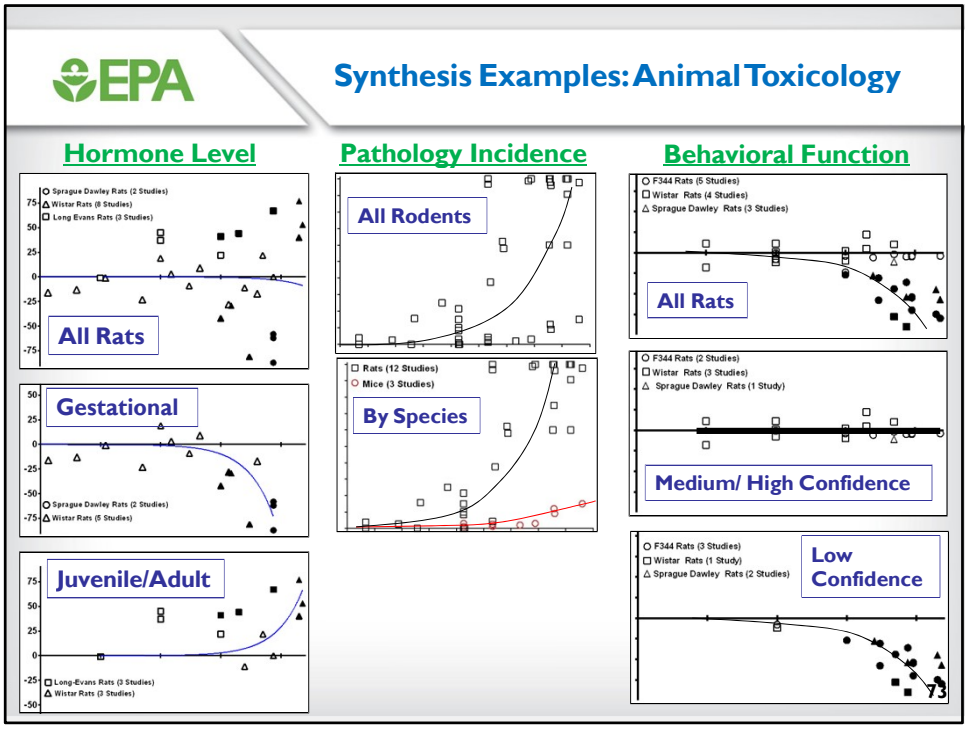
Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation





Appendix C





Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

Mechanistic Evidence

"Mechanistic data represent a wide variety of studies not intended to identify an adverse outcome." (NRC, 2014)

- When evaluating mechanistic evidence, the scope is larger than "*in vitro*" data
- **Mechanistic inventories collected at earlier stages may include:**
  - *In vivo* (cellular, biochemical, molecular)
  - *In vitro* or *ex vivo* (human or animal tissues or cells)
  - Non-animal or non-mammalian alternative animal models
  - Big data ('omics or high-throughput assays)
  - "Intervention" studies (pharmacologic, environmental, genetic)

"…there might be hundreds of *in vitro* and other mechanistic studies of a given chemical…" (NRC, 2014)

"For a given chemical, multiple mechanisms might be involved in a given end point, and it might not be evident how different mechanisms interact in different species to cause the adverse outcome." (NRC, 2014)

74

Systematic review of mechanistic information requires a different approach

"When human data are nonexistent, are mixed, or consistently show no association and an animal study finds a positive association, the importance of mechanistic data is increased..." (NRC, 2014)

To narrow the scope of the analyses of mechanistic information, IRIS applies an iterative approach to identifying key mechanistic questions at various stages of the systematic review

- *Problem formulation* identifies predefined analyses (e.g., when a mutagenic MOA is indicated)
- *Literature inventory* allows identification of studies on an organ system that human and animal studies meeting the PECO criteria have not examined
- *Human and animal evidence syntheses* may flag impactful qualitative and quantitative analyses

75

Human and animal evidence syntheses may flag impactful mechanistic analyses

- Identify precursor events for apical toxicity endpoints
- Inform susceptibility (species, strain, or sex differences; at-risk populations or lifestages)
- Inform human relevance of animal data (note: the level of analysis will vary depending on the impact of the animal evidence)
- Provide biological plausibility (i.e., to human or animal health effect data when evidence is weak or critical uncertainties are identified)
- Establish mechanistic relationships (or lack thereof) across sets of potentially related endpoints/outcomes to inform the consideration of coherence during evidence integration
- Aid extrapolation (high-to-low dose; short-to-long duration; route-to-route)
- Improve dose-response modeling and quantification of uncertainties

76

Mechanistic Analysis Focused on Specific Questions

Examples of when these analyses have been triggered in recent IRIS Assessments:

- Benzo[a]pyrene (2017): The descriptor "carcinogenic to humans" was supported by strong mechanistic evidence that established the biological plausibility of the animal findings occurring in humans, despite lack of human exposure data
  - Key precursors (BPDE-DNA adducts) were identified in humans exposed to PAH mixtures that are specific to B[a]P, form mutational spectra unique to B[a]P, and are associated with cancer in humans
- Dichloromethane (2011): The cancer risk estimate was specifically derived for a susceptible subpopulation (GSTT1+/+) identified by the mechanistic evaluation
  - Differing results *in vivo* were explainable by species and tissue differences in the availability of GST
  - PBPK modeling addressed the variability in this population
- **Documentation and transparency is key for future mechanistic analyses**

77

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

Focused mechanistic evaluations

"Several criteria should be considered in assessing in vitro toxicology studies for risk of bias and toxicologic relevance. Relevance should be determined in several domains, including cell systems used, exposure concentrations, metabolic capacity, and the relationship between a measured in vitro response and a clinically relevant outcome measure. **Few tools are available for assessing risk of bias in in vitro studies. Because of the nascent status of this field, the committee can provide only provisional recommendations for EPA to consider...**EPA should carry out, support, or encourage research on the development and evaluation of empirically based instruments for assessing bias in…mechanistic studies." (NRC, 2014)

- **Prioritize studies of relevant endpoints and associated assays by toxicologic relevance (e.g., model systems; dose range; sensitivity and specificity of assay)**
- **Conduct individual study evaluations on the most impactful studies**
- **EPA is exploring the use of existing tools, including adaptations of IRIS study evaluation tools**
- **Organizational frameworks (e.g., EPA's MOA framework using modified Hill considerations; visual AOP-like constructs) are useful for organizing and documenting these analyses transparently to convey conclusions for evidence integration**

78

Moving from Synthesis to Integration

Outputs of Evidence Synthesis

Results of Human Health Effect Study Synthesis

Results of Animal Health Effect Study Synthesis

Results of Synthesis of Mechanistic Evidence Informing the Human and Animal Syntheses

Evidence Integration

Transparent and Structured Processes for Drawing Summary Conclusions Across Lines of Evidence

79

Appendix C

Evidence Integration Involves a Sequential, Two-Step Process

- **Evidence synthesis interpretations for each consideration relating to causality are combined across lines of evidence using transparent, structured frameworks**

Step 1: "Within-Stream" Integration	Step 2: "Across-Stream" Integration
Judge the Strength of the Evidence from the: • **Human Evidence Stream** • **Animal Evidence Stream**	Draw Overall Evidence Integration Conclusions based on: • Combined **Human** and **Animal Evidence Streams**
Human health effect study synthesis conclusions for each consideration are integrated in light of mechanistic evidence in exposed humans or human cells (or other human models)	The judgments regarding the strength of the human and animal evidence streams are integrated in light of evidence on the human relevance of the findings in animals, susceptibility, and the coherence of the findings across evidence streams.
Characterize the Strength of the Evidence for an Effect in Animals (**Animal Evidence Stream Judgment**)	
Animal health effect study synthesis conclusions for each consideration are integrated in light of mechanistic evidence in exposed animals or animal cells (or other relevant models)	

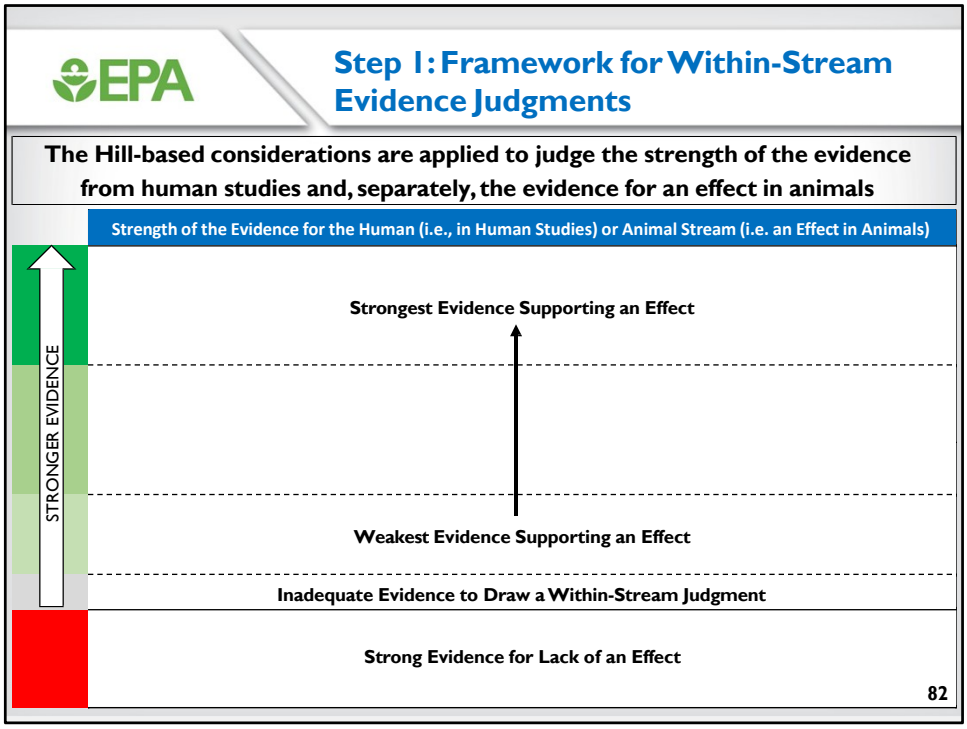
80

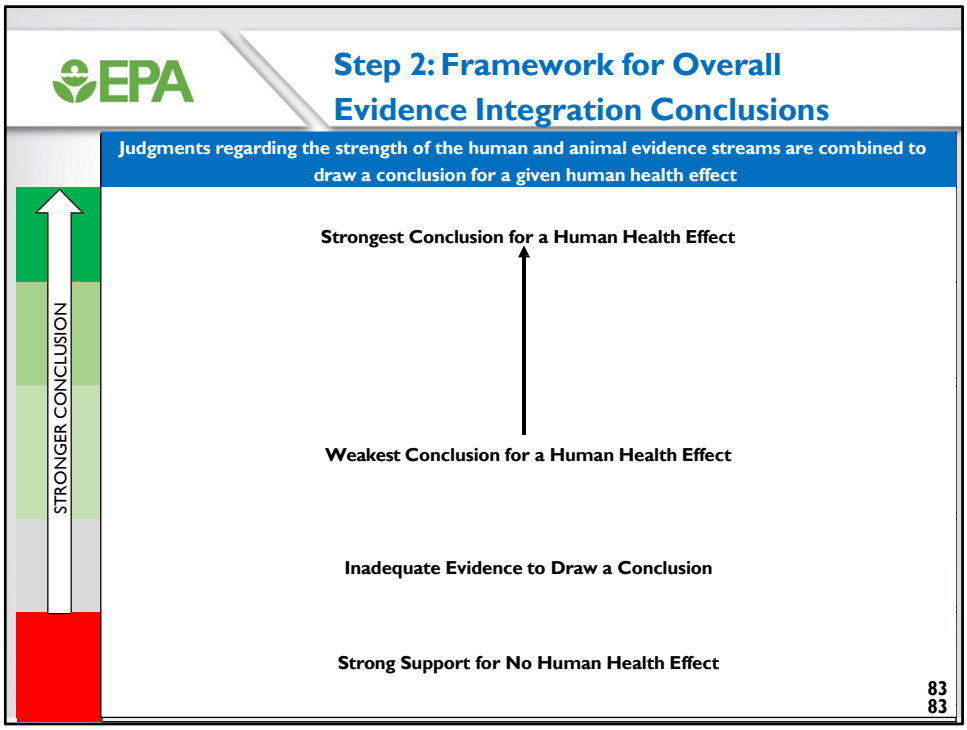
Within-Stream (Human; Animal Stream) Evidence Judgment Considerations

	Human Evidence Stream	Animal Evidence Stream
Individual Studies	• *High* or *medium* confidence studies provide stronger evidence within evaluations of each Hill consideration • Interpreting results considers biological as well as statistical significance, and findings across studies	
Consistency	• Different studies or populations increase strength	• Different studies, species, or labs increase strength
Dose-response	• Simple or complex (nonlinear) relationships provide stronger evidence • Dose-dependence that is expected, but missing, can weaken evidence (after considering the findings in the context of other available studies and biological understanding)	
Magnitude, Precision	• Large or severe effects can increase strength; further consider imprecise findings (e.g., across studies) • Small changes don't necessarily reduce evidence strength (consider variability, historical data, and bias)	
Coherence	• Biologically related findings within an organ system, within or across studies, or across populations (e.g., sex) increases evidence strength (considering the temporal- and dose-dependence of the relationship) • An observed lack of expected changes reduces evidence strength	
	• Informed by mechanistic evidence on the biological development of the health effect or toxicokinetic/ dynamic knowledge of the chemical or related chemicals	
Mechanistic Evidence on Biological Plausibility	• Mechanistic evidence in humans or animals of precursors or biomarkers of health effects, or of changes in established biological pathways or a theoretical mode-of-action, can strengthen evidence • Lack of mechanistic understanding does not weaken evidence outright, but it can if well-conducted experiments exist and demonstrate that effects are unlikely	

Light blue rows highlight mechanistic inferences; "temporality" and "natural experiments" not shown

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation





Appendix C

Step 2: Framework for Overall Evidence Integration Conclusions

Judgments regarding the strength of the human and animal evidence streams are combined to draw a conclusion for a given human health effect

STRONGER CONCLUSION

A very high level of certainty that exposure causes the health effect in humans, e.g.,:
- The strongest evidence judgment for the human evidence stream
- A moderately strong human evidence judgment and the strongest animal evidence judgment alongside strong mechanistic evidence that MOAs and key precursors identified in animals are anticipated to occur in humans

Reasonable certainty that exposure causes the health effect in humans, although some outstanding questions may remain, e.g.,:
- The strongest evidence judgment for the animal evidence stream, but not meeting the criterion above
- A moderately strong human or animal evidence stream judgment, or the weaker judgments when evidence from the opposite stream (e.g., mechanistic evidence of precursors supporting coherence) that increases certainty

Conveys some concern that exposure may cause a particular health outcome in humans, but either there were very few studies that contributed to the evaluation, the evidence was weak or conflicting, and/ or the methodological conduct of the studies was poor. Given the substantial degree of uncertainty, additional research is encouraged. Scenarios include:
- The weakest human or animal evidence stream judgment, or a moderately strong judgment with evidence from the opposite stream (e.g., null results in well-conducted mechanistic studies of precursors) that decreases certainty
- Exceptionally, strong mechanistic evidence in the absence of conventional human or animal studies

This conveys either a lack of information or an inability to interpret the available evidence, e.g.,:
- Inadequate evidence to judge the strength of both the human and animal evidence streams.
- The strongest animal evidence stream judgment with inadequate evidence to judge the strength of the human evidence, and with strong mechanistic information indicating that the animal evidence is unlikely to be relevant to humans.

A substantial degree of certainty that there is negligible concern for exposure to cause the health effect in humans, e.g.,:
- Meeting the criteria for drawing a judgment of 'strong support for no effect' for the human evidence stream
- Meeting the criteria for drawing a judgment of 'strong support for no effect' for the animal evidence stream along with inadequate evidence to judge the strength of the human evidence and strong mechanistic support that the animal models are able to identify an association

84

Evidence Profile Table: Supports the Evidence Integration Narrative

"the weight of evidence descriptions need to indicate the various determinants of weight... to be able to understand what elements (such as consistency) were emphasized" [NRC, 2011]; "No matter what method is used to integrate the different kinds of evidence available for an IRIS assessment, using a template for the evidence-integration narrative could help to make IRIS assessments more transparent." [NRC, 2014]

Studies and interpretation	Factors that increase strength	Factors that decrease strength	Summary of findings	Within stream evidence judgments	Inference across evidence streams	Overall conclusion
[Health Effect or Outcome Grouping]						
Evidence from Human Studies (Route) • *References* • *Study confidence (based on evaluation of risk of bias and sensitivity) and explanation* • *Study design description*	• *Consistency* • *Dose-response gradient* • *Coherence of observed effects (apical studies)* • *Effect size (magnitude, severity)* • *Biological plausibility* • *Low risk of bias/ high quality* • *Insensitivity of null/ negative studies* • *Natural experiments* • *Temporality*	• *Unexplained inconsistency* • *Imprecision* • *Indirectness/ applicability* • *Poor study quality/ high risk of bias* • *Other (e.g., Single/Few Studies; small sample size)* • *Evidence demonstrating implausibility*	• *Results information (general endpoints affected/ unaffected) across studies* • *Human mechanistic evidence informing biological plausibility: discuss how data influenced the within stream judgment (e.g., evidence of precursors in exposed humans).* *Could be multiple rows (e.g., grouped by study confidence or population) if this informs results heterogeneity*	*Describe strength of the evidence from human studies, and primary basis:* +++ Strongest evidence ++◯ +◯◯ Weakest evidence ◯◯◯ – ◯◯◯ – ◯◯◯ Inadequate – – – Strong evidence for no effect	*Human relevance of findings in animals* • *Cross-stream coherence (i.e. for both health effect-specific and mechanistic data)* • *Other inferences:* ○ *Information on susceptibility* ○ *MOA analysis inferences: precursors, cross-species inferences of toxicokinetics, or quantitative implications* ○ *Relevant information from other sources (e.g., read across; other, potentially related health hazards)*	*Describe conclusion(s) and primary basis for the integration of all available evidence (across human, animal, and mechanistic):* +++ Strongest conclusion ++◯ +◯◯ Weakest conclusion ◯◯◯ – ◯◯◯ – ◯◯◯ Inadequate – – – Strong support for no human health effect Summarize the models and range of dose levels upon which the conclusions were primarily reliant
Evidence for an Effect in Animals (Route) • *References* • *Study confidence (based on evaluation of risk of bias and sensitivity) and explanation* • *Study design description*	• *Consistency and Replication* • *Dose-response gradient* • *Coherence of observed effects (apical studies)* • *Effect size (magnitude, severity)* • *Biological plausibility* • *Low risk of bias/ high quality* • *Insensitivity of null/ negative studies*	• *Unexplained inconsistency* • *Imprecision* • *Indirectness/ applicability* • *Poor study quality/ high risk of bias* • *Other (e.g., Single/Few Studies; small sample size)* • *Evidence demonstrating implausibility*	• *Results information (general endpoints affected/ unaffected) across studies* • *Animal mechanistic evidence informing biological plausibility for effects in animals: discuss how mechanistic data influenced the within stream judgment (e.g., evidence of coherent molecular changes in animal studies)* *Could be multiple rows (e.g., by study confidence, species, or exposure duration) if this informs results heterogeneity*	*Describe strength of the evidence for an effect in animals, and primary basis:* +++ Strongest evidence ++◯ +◯◯ Weakest evidence ◯◯◯ – ◯◯◯ – ◯◯◯ Inadequate – – – Strong evidence for no effect		

85

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

Evidence Integration Conclusions

- **For Cancer**, conclusions on the integrated evidence for each cancer type (or grouping) are evaluated in the context of MOA information to develop an evidence integration narrative that includes a descriptor for carcinogenicity:
  - **carcinogenic** to humans; **likely** to be carcinogenic to humans; **suggestive** evidence of carcinogenic potential; **inadequate** information to assess carcinogenic potential; or **not likely** to be carcinogenic to humans
- **For Noncancer Effects**, frameworks for evaluating the integrated evidence have been developed to add structure and transparency to the evidence integration narrative(s), which include(s) the relevant exposure context.
  - IRIS has not yet incorporated standardized descriptors for noncancer effects
  - The NAS recommended incremental improvements in this area, including recommendations to "Develop uniform language to describe strength of evidence on noncancer effects" [p. 92, 2014]
  - The specific way in which these conclusions are summarized is currently being tested and discussed within EPA

86

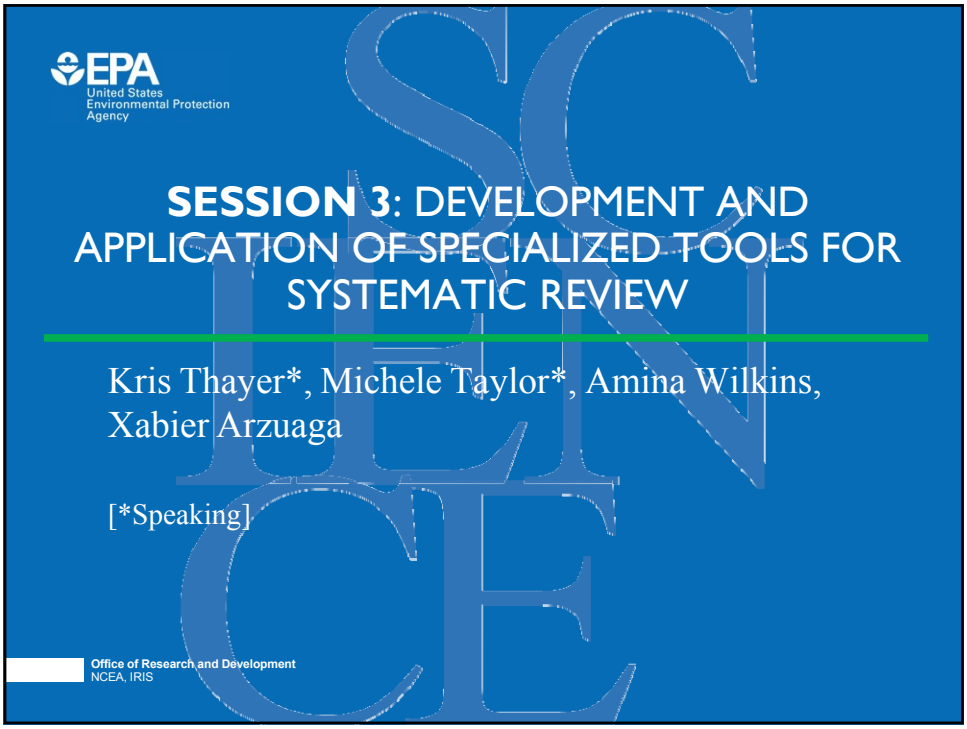
IRIS has Addressed the Major NAS 2014 Recommendations

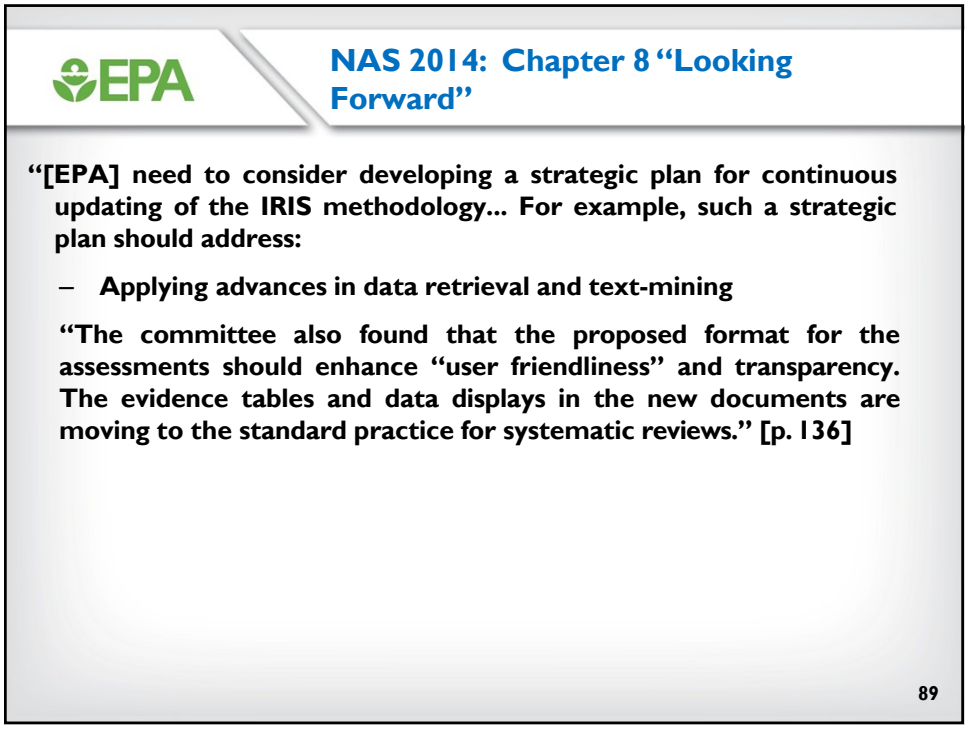
NAS 2014 Topics	IRIS Process Improvements
Evidence Evaluation (Chapter 5)	• Individual studies are evaluated for reporting quality, risk of bias, and sensitivity • Decisions and supporting rationale are clearly documented • Study evaluations impact subsequent assessment decisions
Evidence Integration for Hazard Identification (Chapter 6)	• Structured frameworks provide transparency in expert judgments across human, animal, and mechanistic studies (based on Hill) • Standardized templates documenting key evidence integration decisions have been developed (evidence profile tables)

See Posters and Demonstrations:

- Male reproductive toxicity in studies of phthalates (4 posters on a case study for each of the 3 lines of evidence and the overall evidence integration)
- Combining data within species (poster on meta-analytical approaches)
- PBPK model evaluation for human health assessments (poster)
- Health Assessment Workspace Collaborative (demonstration)

87





Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

Current Application of Systematic Review Software

- Specialized software tools make the process more efficient
  - Time and cost savings, improved data management, increased transparency
- NOT all systematic review software tools are intended to automate/semi-automate the process, e.g., HAWC helps manage information content
  - Currently, automation tools are most advanced for evidence identification
- Prefer free tools when possible to help address needs of a potentially large community of users in environmental and biomedical sciences
- Incorporate tools after confirming acceptable performance and interoperability with HERO
  - A toolbox approach, not a "one and only" tool model
- Organized multiple IRIS staff training sessions in 2017 and created a support team ("train the trainers" model)

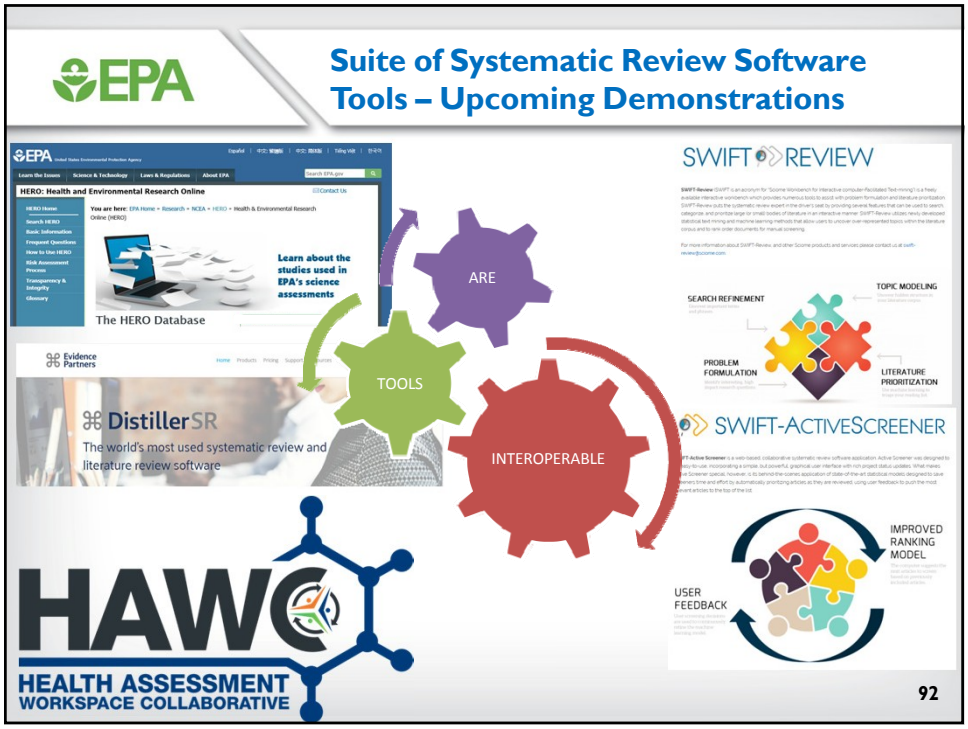
90

Research Activities

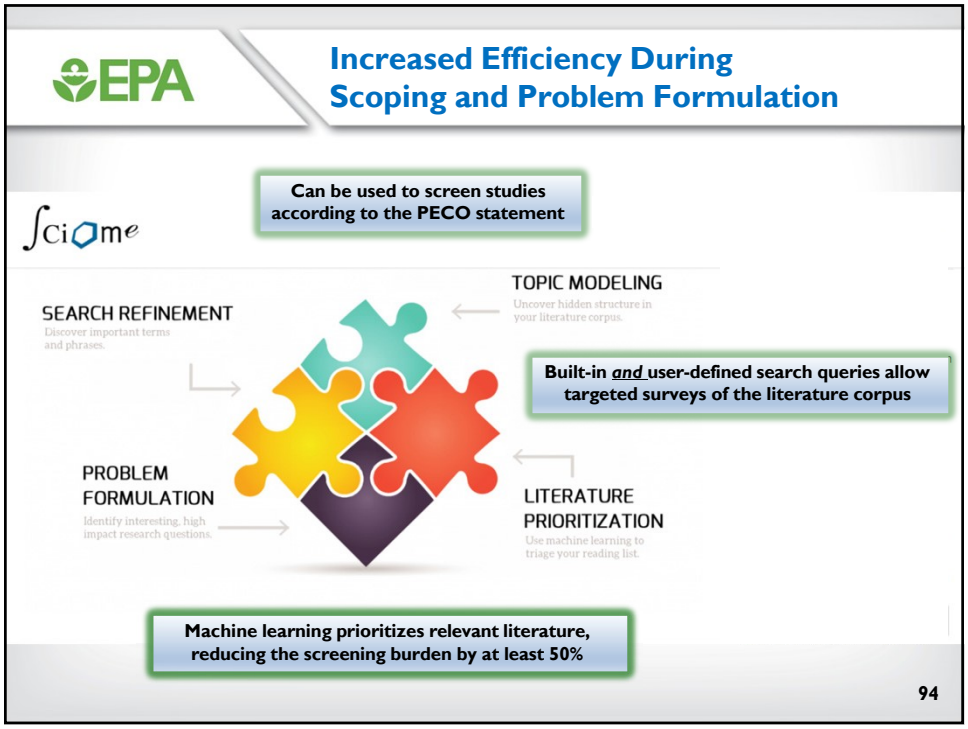
- Developing tools to help automate beyond evidence identification is a long-term research commitment
  - Major hurdle is lack of training/test sets for model development
  - Better performance expected for more structured content (e.g., animal bioassay compared to epidemiological studies)
- Any progress on semi-automation could result in large time and cost savings
- In 2017, NCEA created an interagency agreement with NTP to leverage resources
  - Current activities focus on creating test/training sets and model development for basic content of animal studies (e.g., test chemical, species, dose levels, randomization, etc.).
  - Other parts of EPA can also utilize interagency agreement
- Innovation challenges may be required to identify solutions for capturing complex content, i.e., table content, information spread across multiple sentences and paragraphs

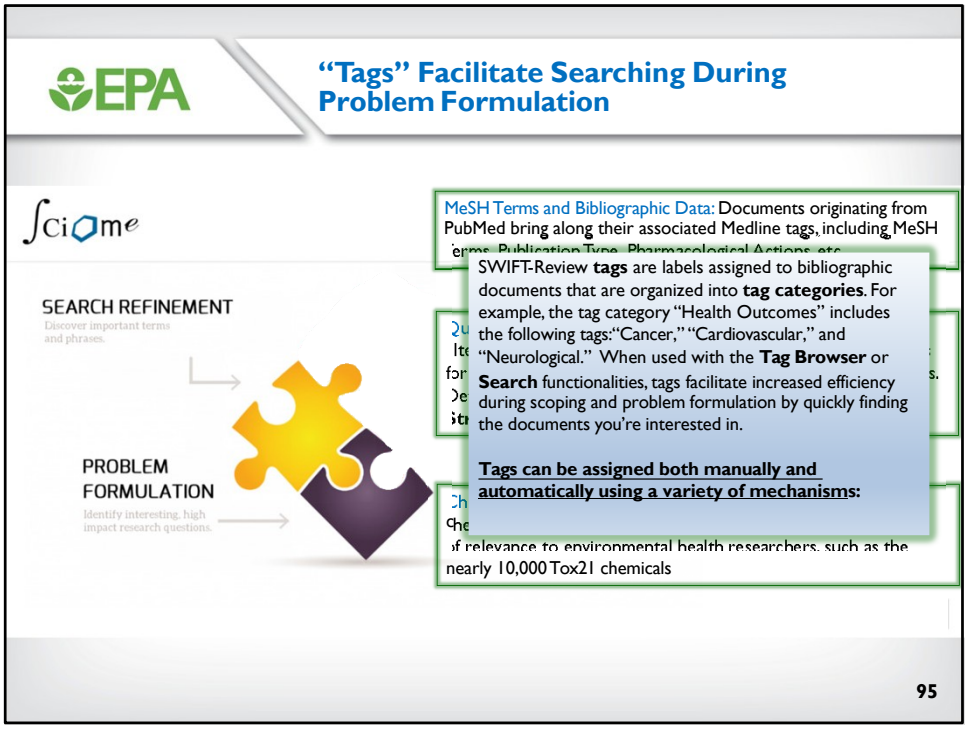
91

Appendix C

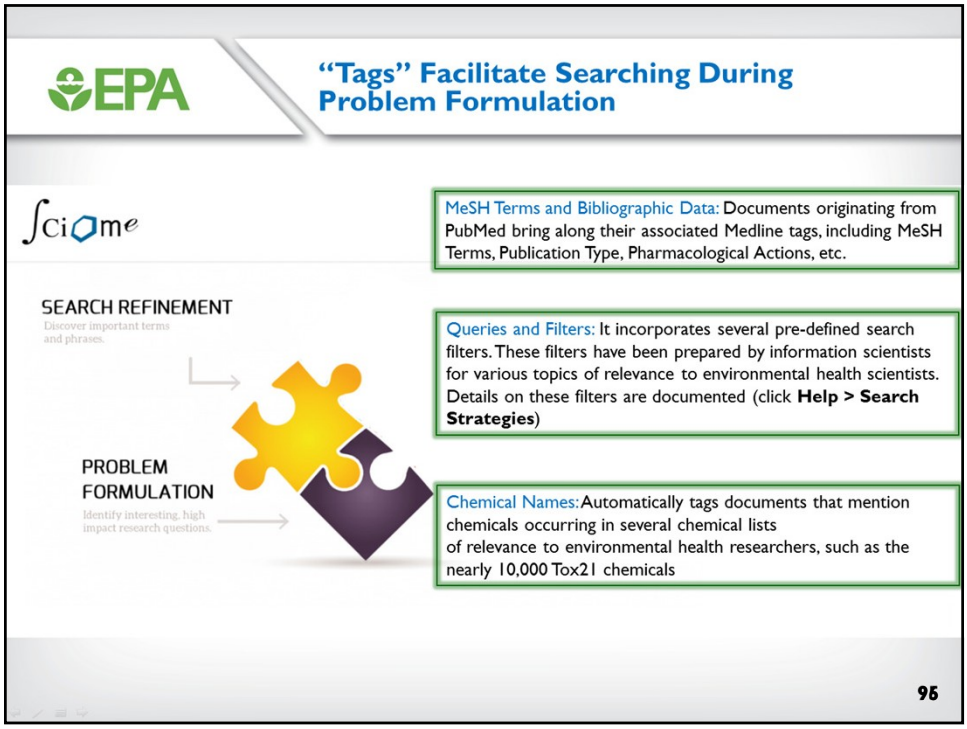


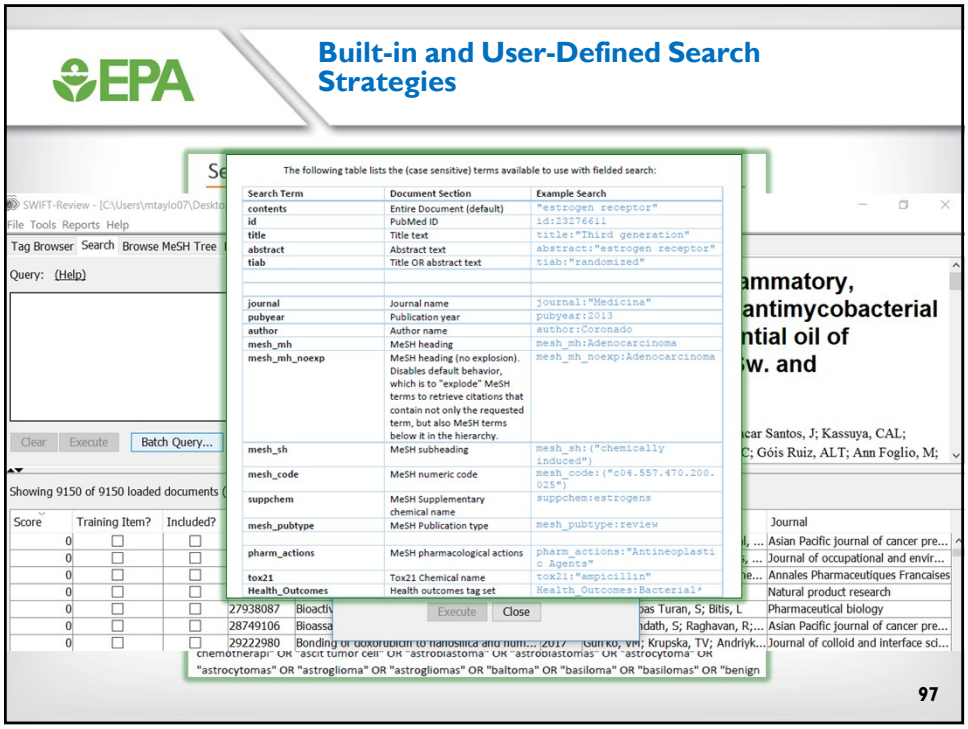
Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation



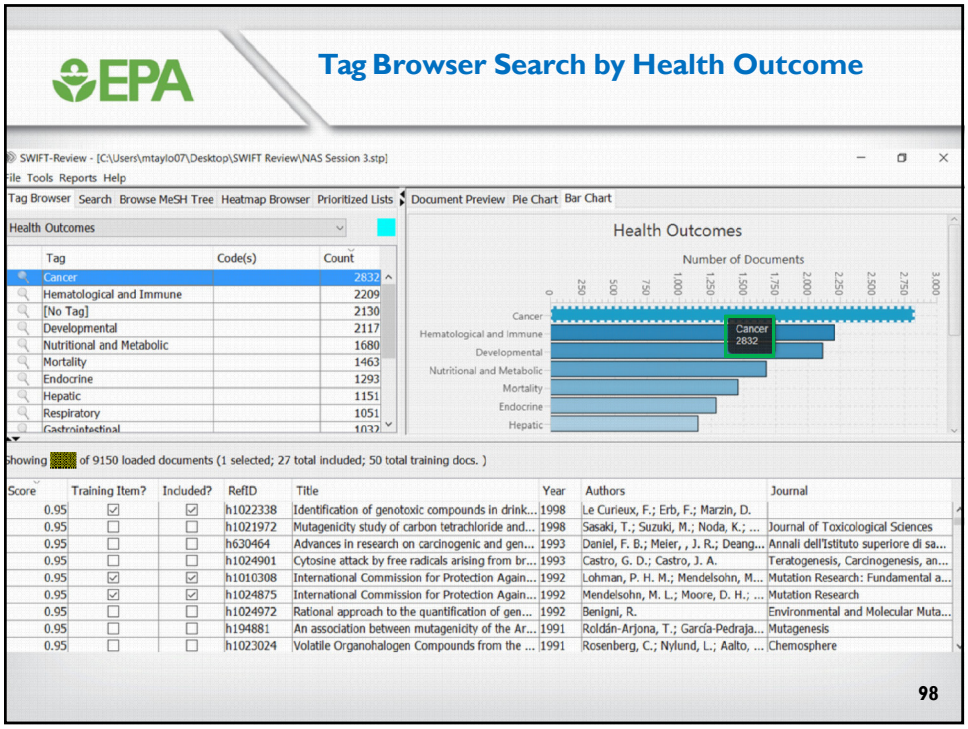


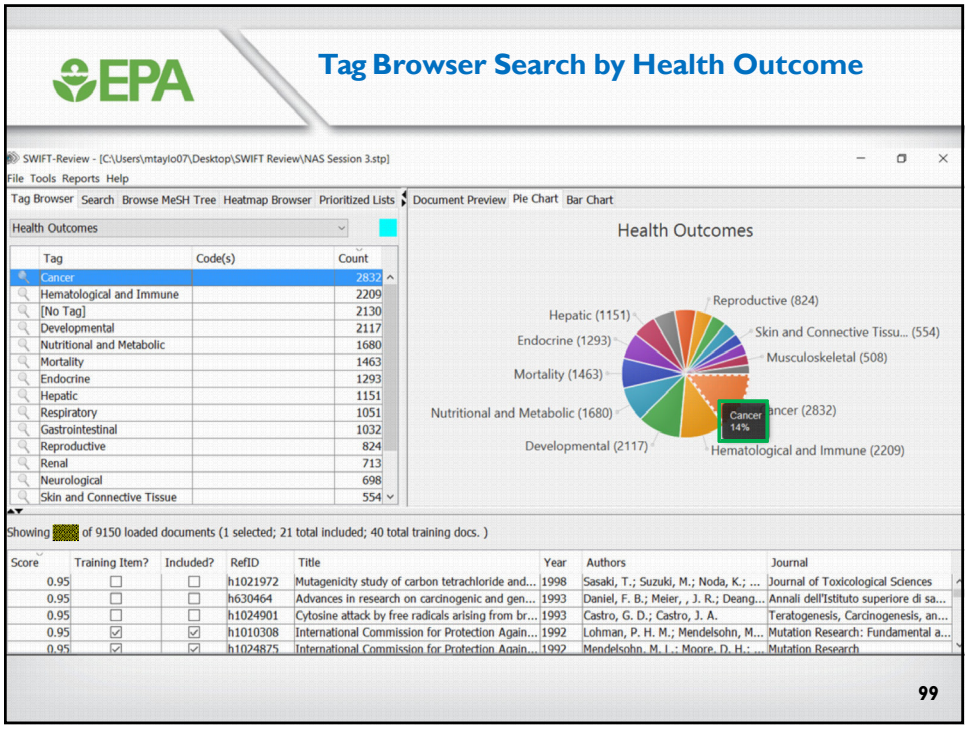
Appendix C



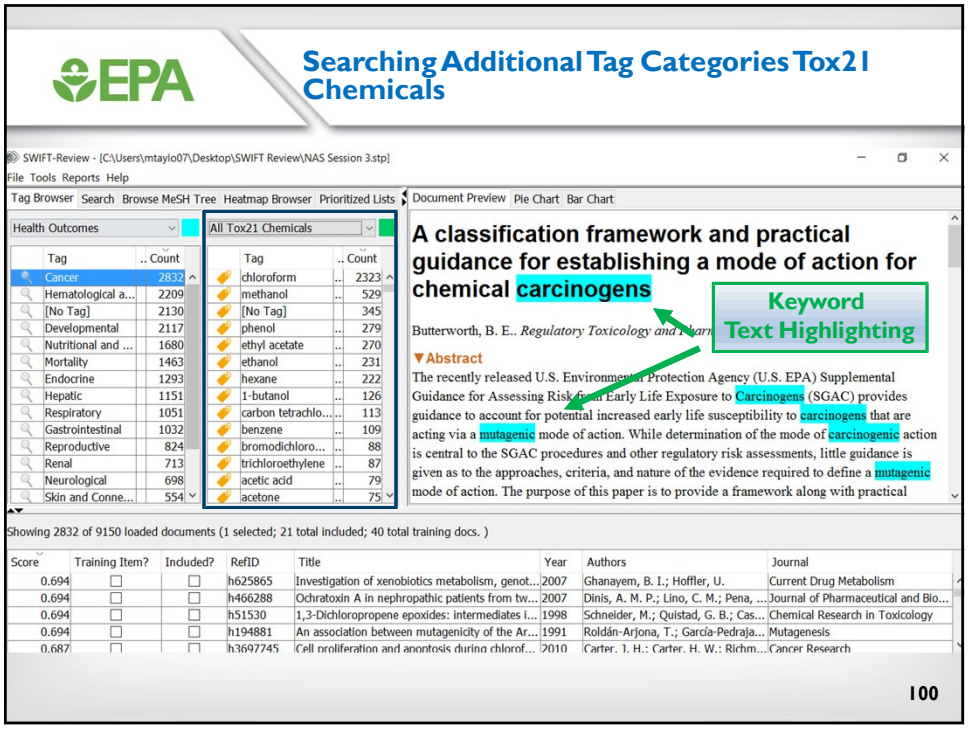


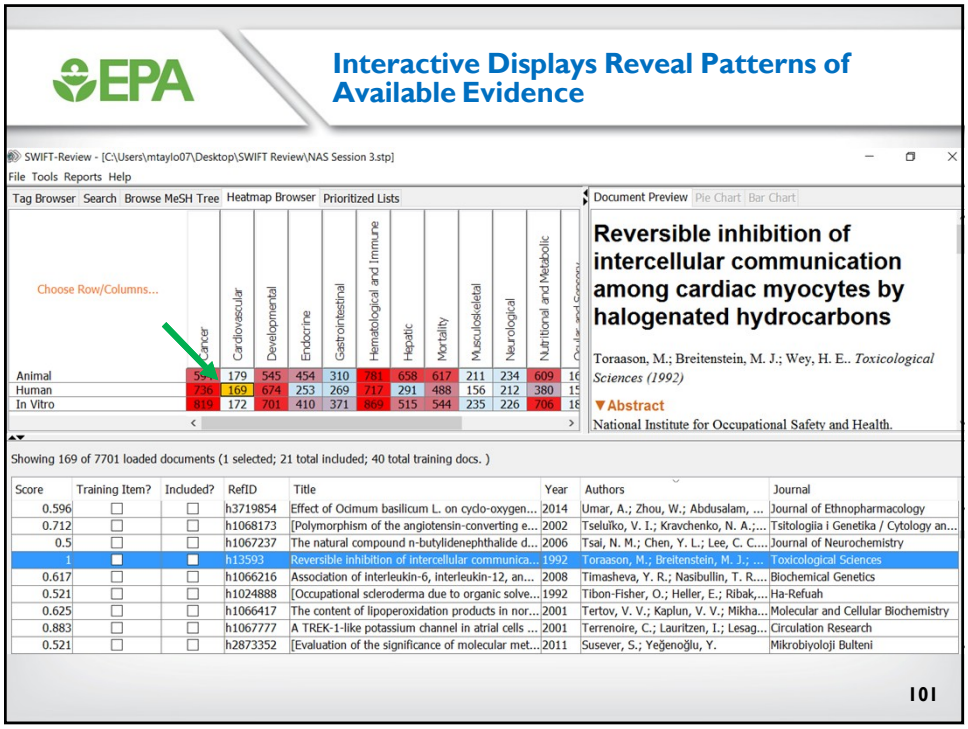
Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation



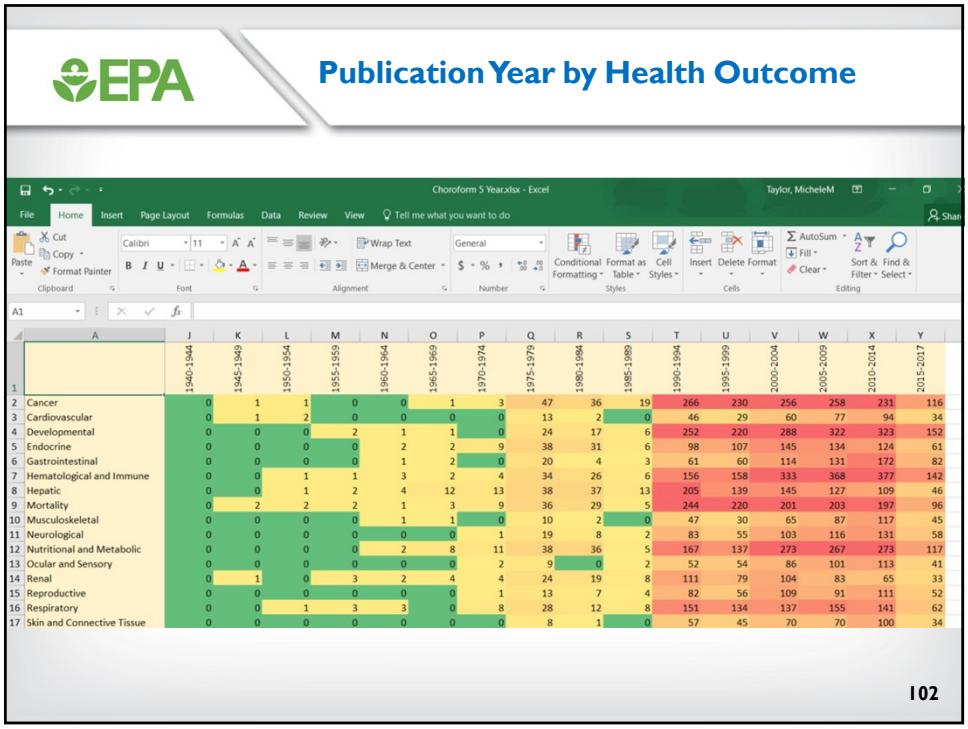


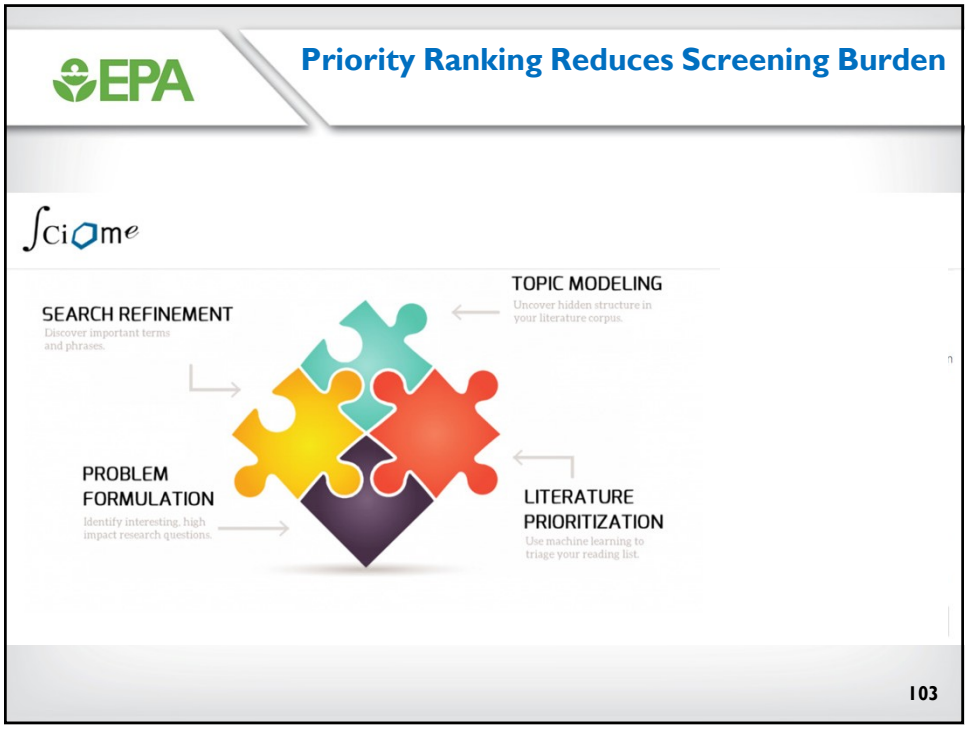
Appendix C



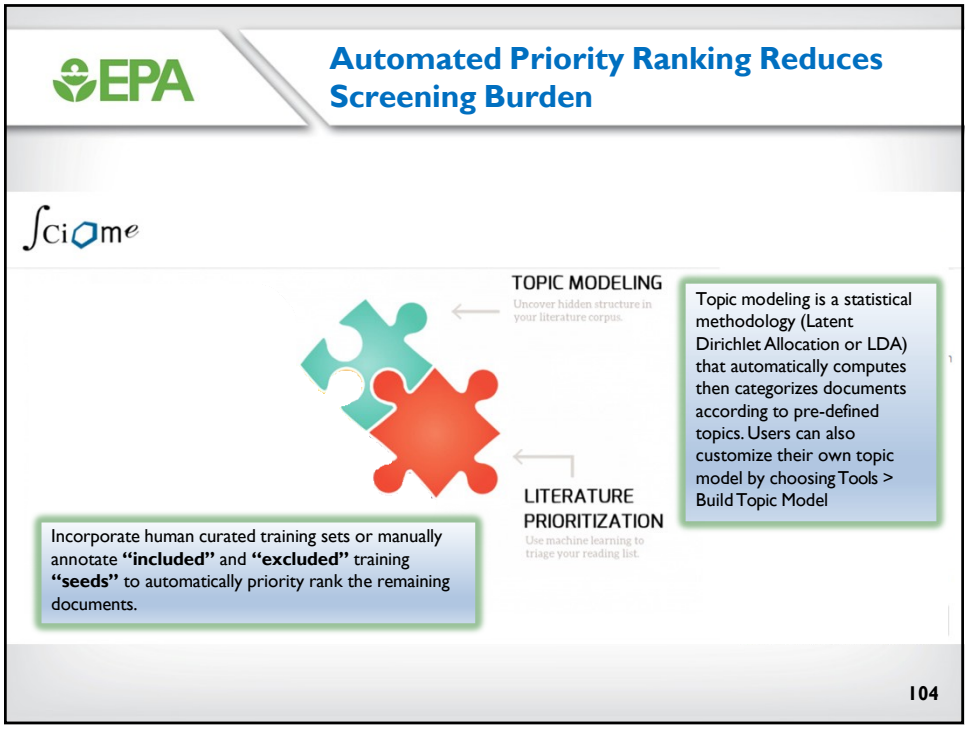


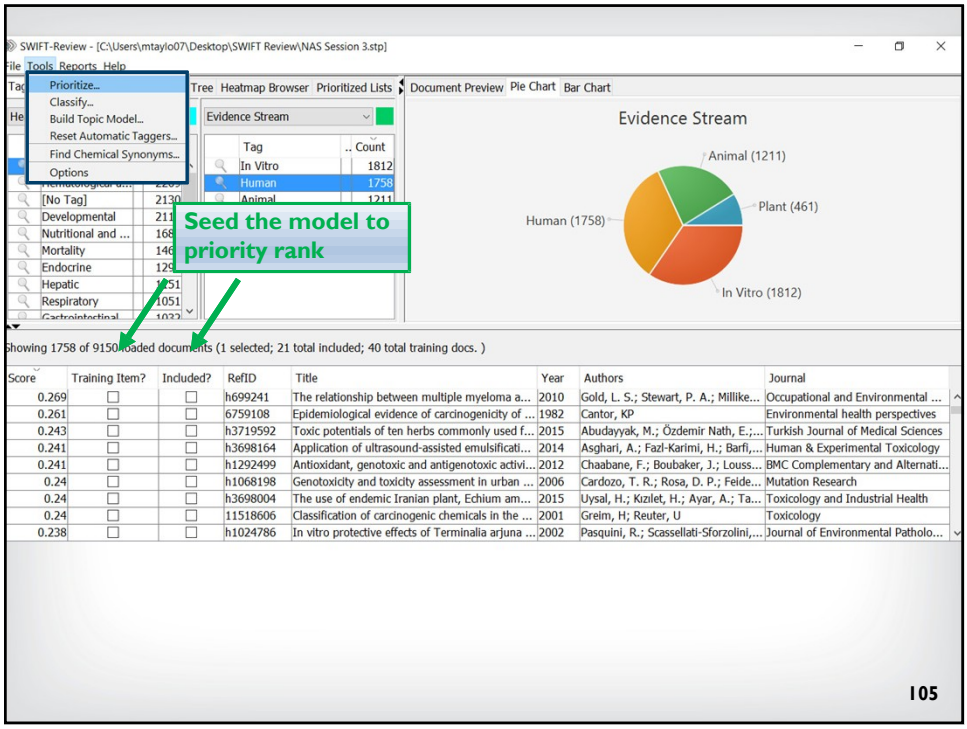
Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation



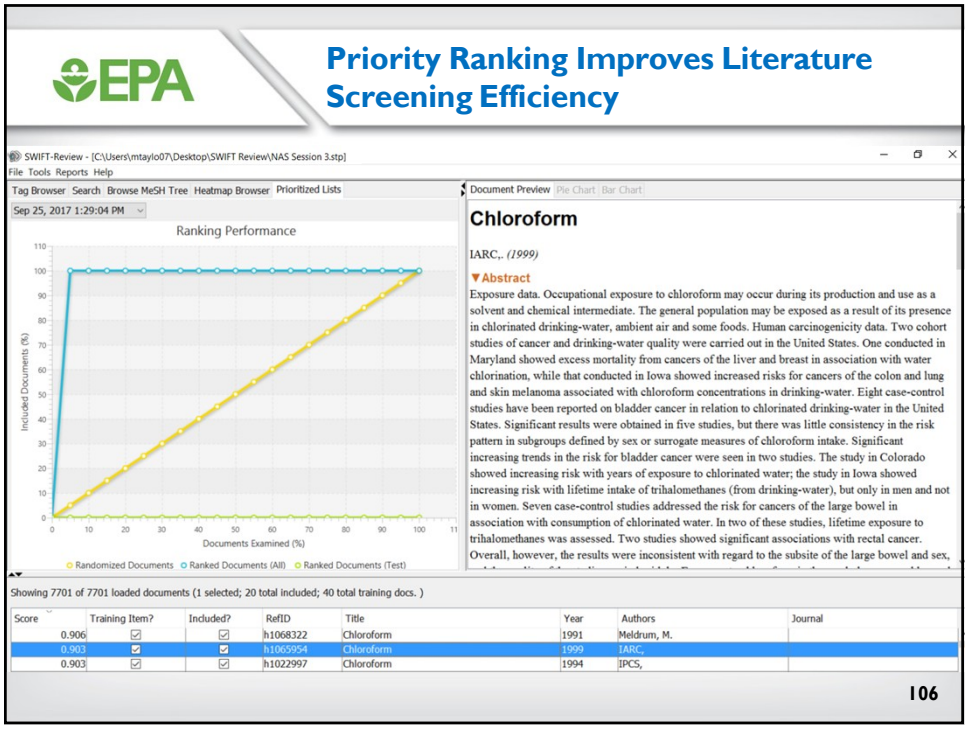


Appendix C





Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

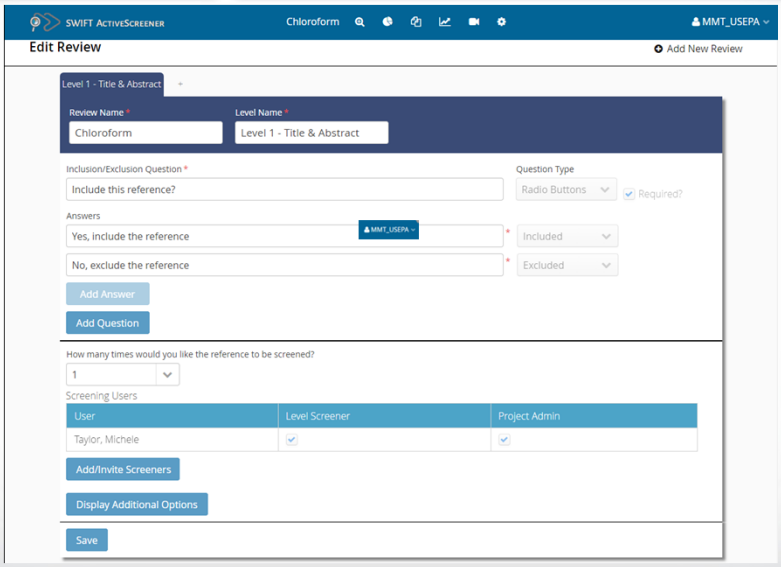


SWIFT Active Screener Capabilities - Improved Ranking Model

- **Web-based, real-time, collaborative, systematic review software application**
- **State-of-the-art statistical models prioritize articles *as they are being reviewed***
- **Experience suggests screening burden is reduced by at least 50% (likely more)**
- **Algorithm improves from screener-input without training "seeds" further increasing efficiency (more efficient than implementing a "seed studies" only model)**
- **Option to "seed" studies if relevant on/off topic literature has been identified**
- **Incorporates a graphical user interface to provide project status updates**
- **User-defined screening levels**
  - **Level 1: Title and Abstract**
  - **Level 2: Full text screening**
  - **Level 3: Conflict Resolution**

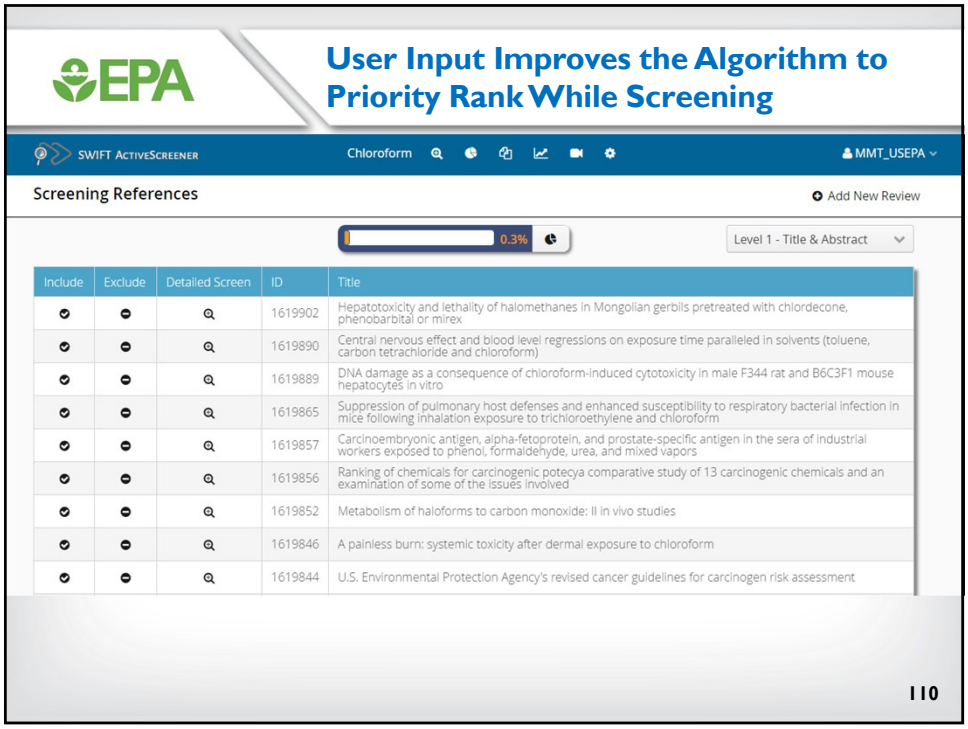
108

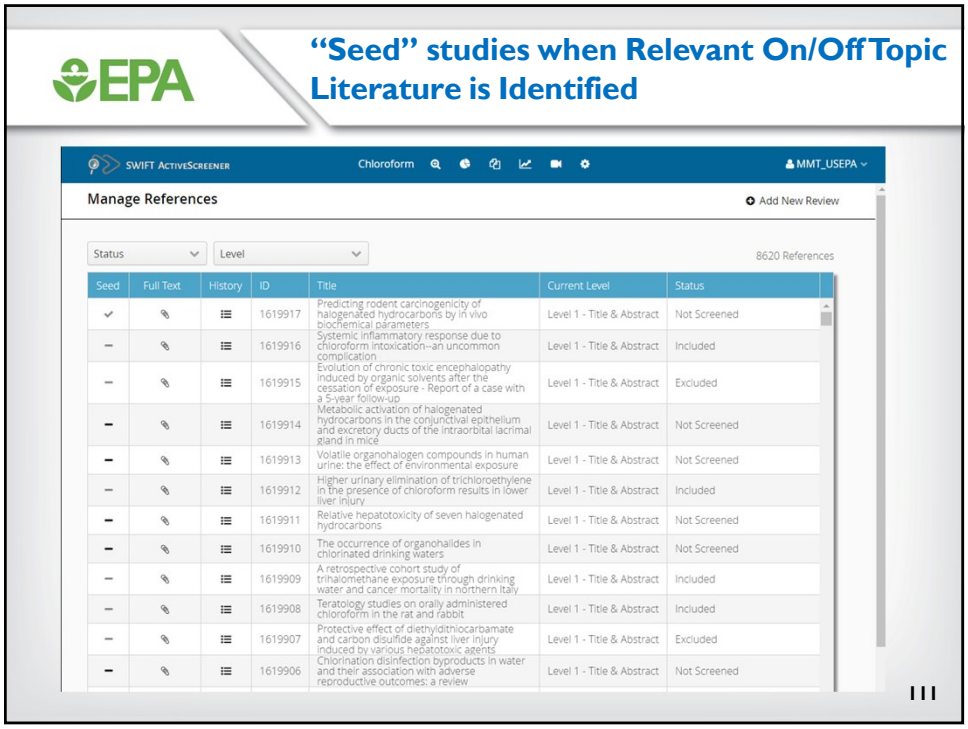
Customize Inclusion/Exclusion Criteria According to the PECO Statement



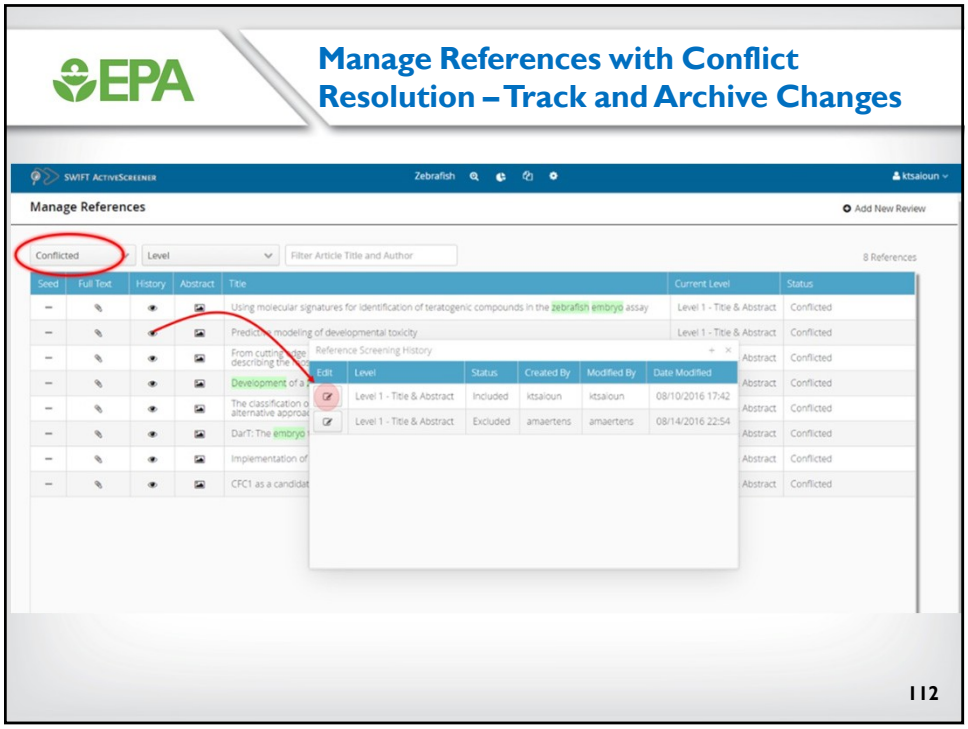
109

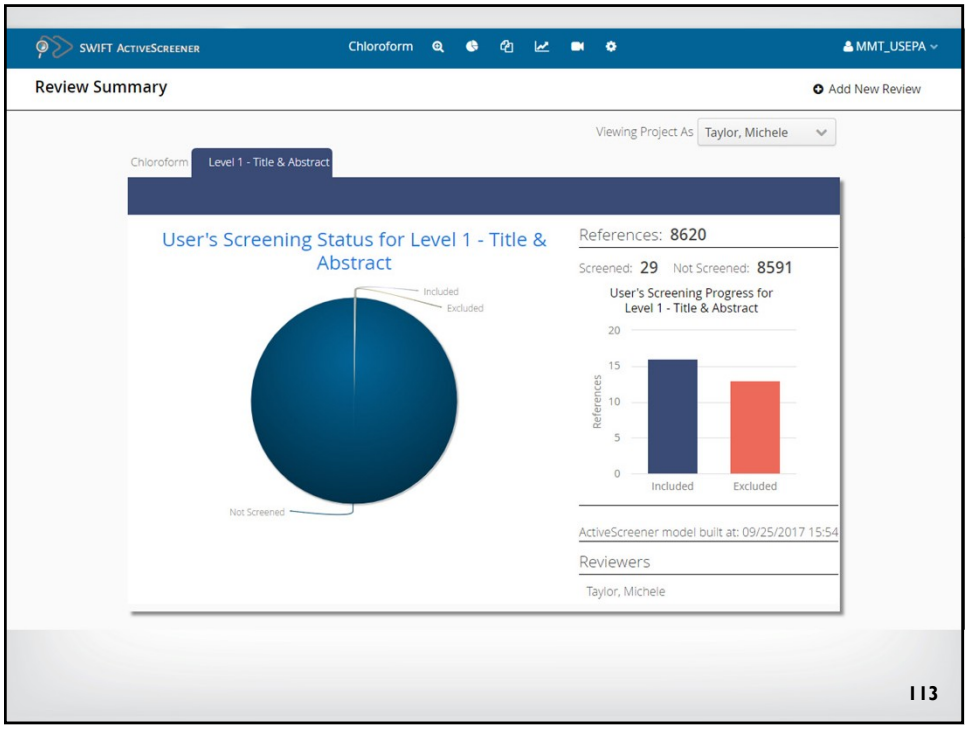
Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation





Appendix C





Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

SWIFT Active: Data Integration

- **Active Screener integrates with systematic review tools already in use:**
  - Accepts imports from bibliographic databases and reference curation platforms including *SWIFT Review*, EndNote, Mendeley, Zotaro, and PubMed
  - Results from screening in Active Screener can be exported in standard data formats compatible with applications including *HAWC* and Excel, EndNote, Mendeley, and Zotaro

Current Users

NIH NIEHS — National Institute of Environmental Health Sciences (NIEHS)
United States Environmental Protection Agency (EPA)
TEDX The Endocrine Disruption Exchange — The Endocrine Disruption Exchange (TEDX)
ebtc — Johns Hopkins Evidence-Based Toxicology Collaboration (EBTC)
USDA — United States Department of Agriculture (USDA)

114

HAWC: Study Evaluation, Extraction, Visualization and Data Sharing

HAWC
HEALTH ASSESSMENT WORKSPACE COLLABORATIVE
https://hawcproject.org/

115

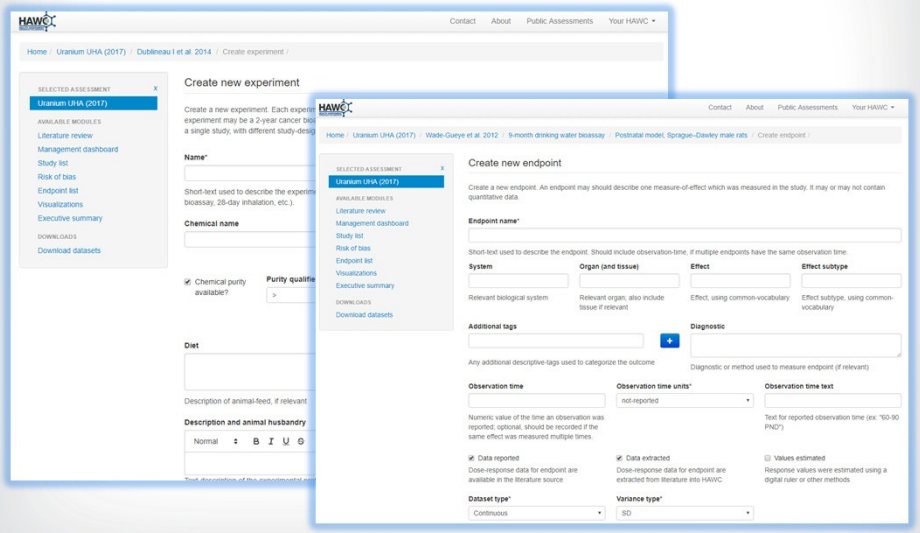
HAWC Capabilities

- **Free and open source**
- **Developed at UNC by Andy Shapiro* with Ivan Rusyn**
- **Literature search and initial screening**
- **Animal bioassay, epidemiological, and in vitro structured study methods/data extraction and visualization**
- **Interactive "click to see more" graphics**
- **Risk of bias and sensitivity evaluation**
- **Modular to work with other tools and maximize flexibility for users**
- **Works best in Google Chrome (preferred), Mozilla Firefox, and Safari**

*current affiliation is National Institute of Environmental Health Sciences/National Toxicology Program (NIEHS/NTP)

116

HAWC: Summarizing Animal Bioassays

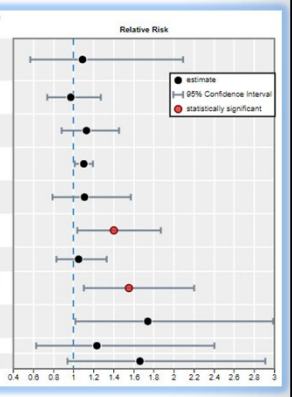


117

Epidemiology: Click to See More Display

Example from Chloroform

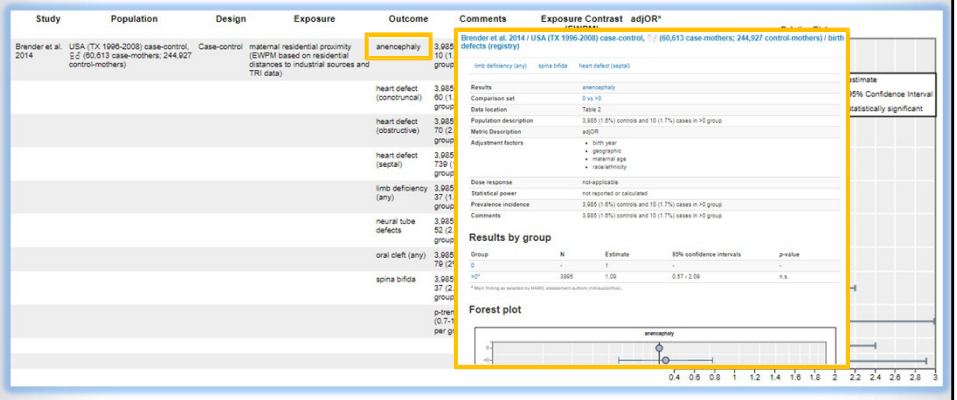
Study	Population	Design	Exposure	Outcome	Comments	Exposure Contrast (EWPM)	adjOR*
Brender et al. 2014	USA (TX 1996-2008) case-control, [illegible] (60,613 case-mothers; 244,927 control-mothers)	Case-control	maternal residential proximity (EWPM based on residential distances to industrial sources and TRI data)	anencephaly	3,985 (1.6%) controls and 10 (1.7%) cases in >0 group	0 vs >0	1.09
				heart defect (conotruncal)	3,985 (1.6%) controls and 60 (1.8%) cases in >0 group	0 vs >0	0.97
				heart defect (obstructive)	3,985 (1.6%) controls and 70 (2.1%) cases in >0 group	0 vs >0	1.13
				heart defect (septal)	3,985 (1.6%) controls and 739 (1.7%) cases in >0 group	0 vs >0	1.1
				limb deficiency (any)	3,985 (1.6%) controls and 37 (1.9%) cases in >0 group	0 vs >0	1.11
				neural tube defects	3,985 (1.6%) controls and 52 (2.4%) cases in >0 group	0 vs >0	1.4
				oral cleft (any)	3,985 (1.6%) controls and 79 (2%) cases in >0 group	0 vs >0	1.05
				spina bifida	3,985 (1.6%) controls and 37 (2.8%) cases in >0 group	0 vs >0	1.55
					p-trend = 0.027; 9-14 (0.7-1.1%) exposed cases per group	0 vs 0.01-42.27	1.74
						0 vs 42.28-1490.26	1.23
						0 vs >1490.26	1.66



118

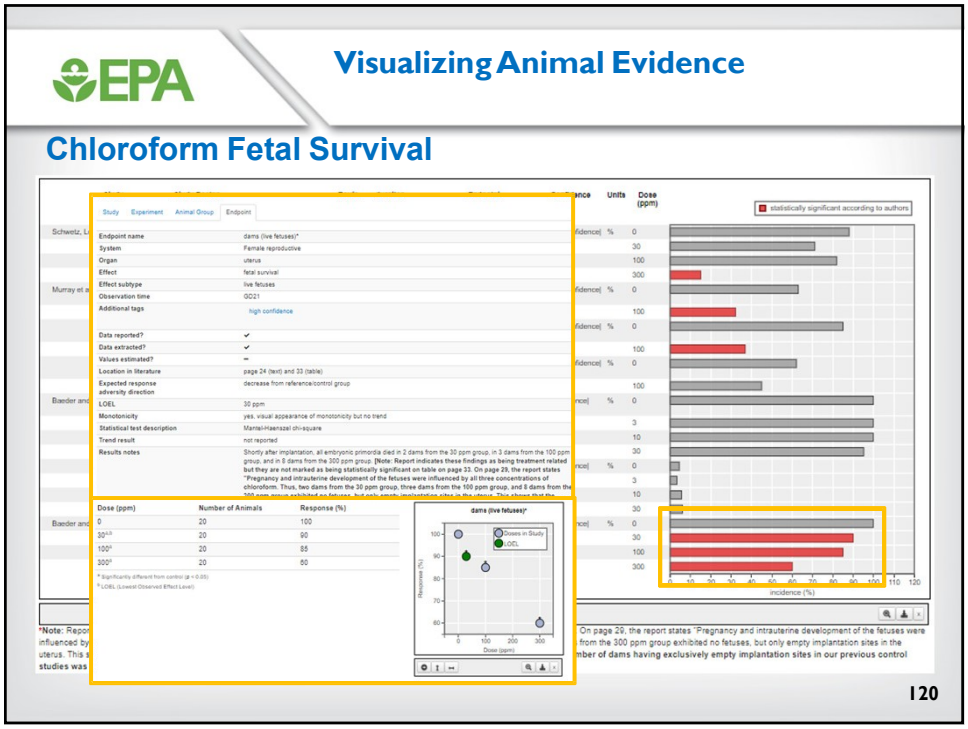
Visualizing Epidemiology Evidence

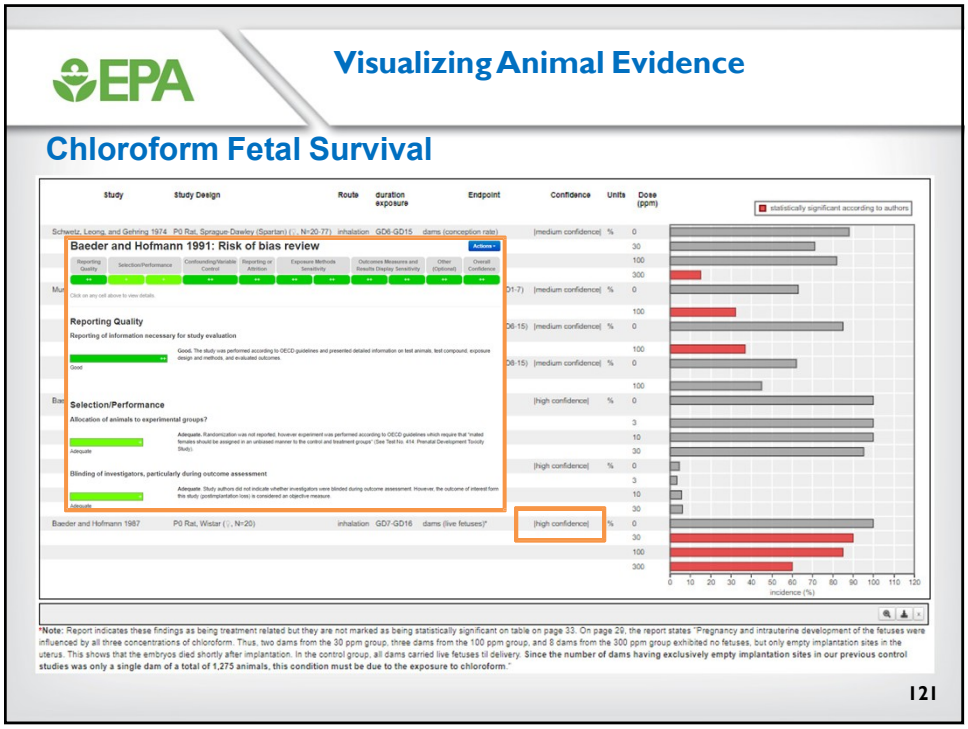
Example from Chloroform

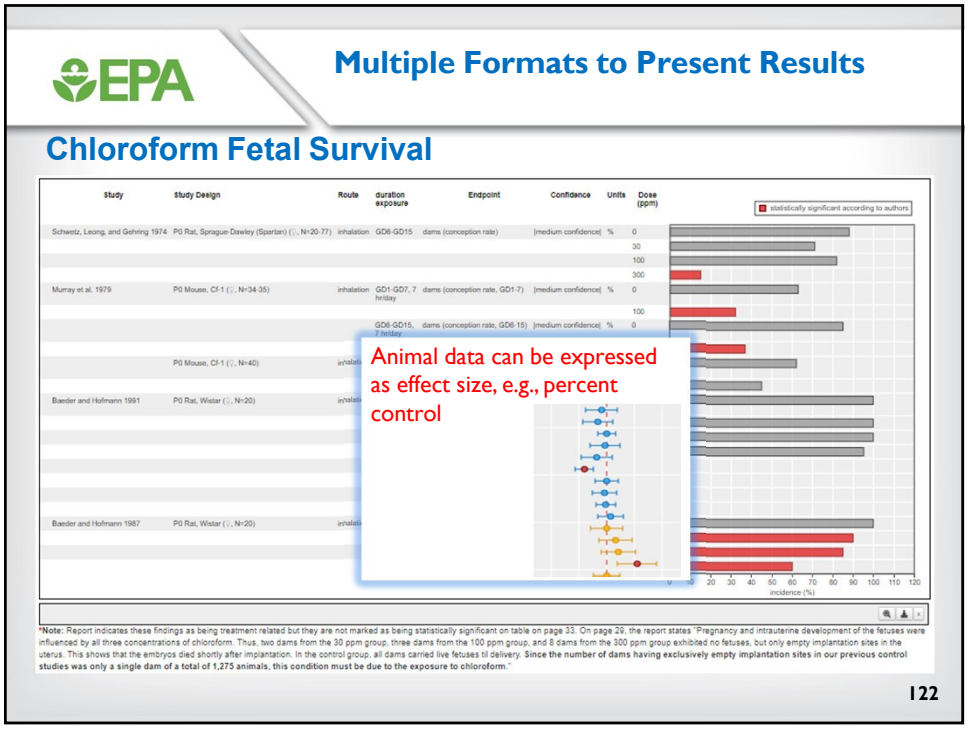


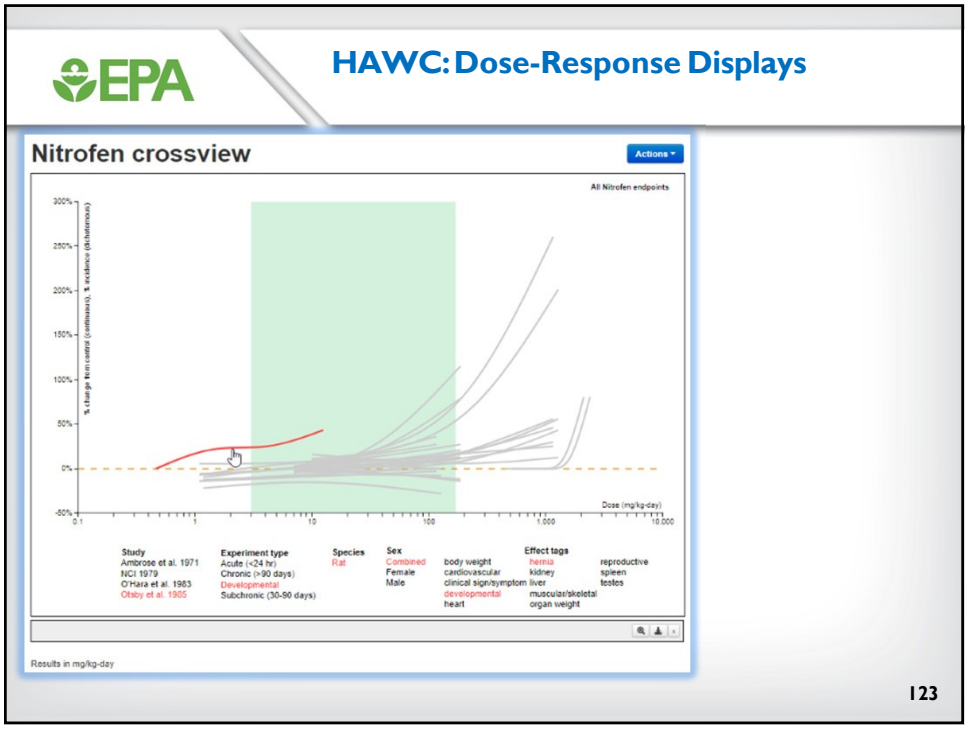
119

Appendix C

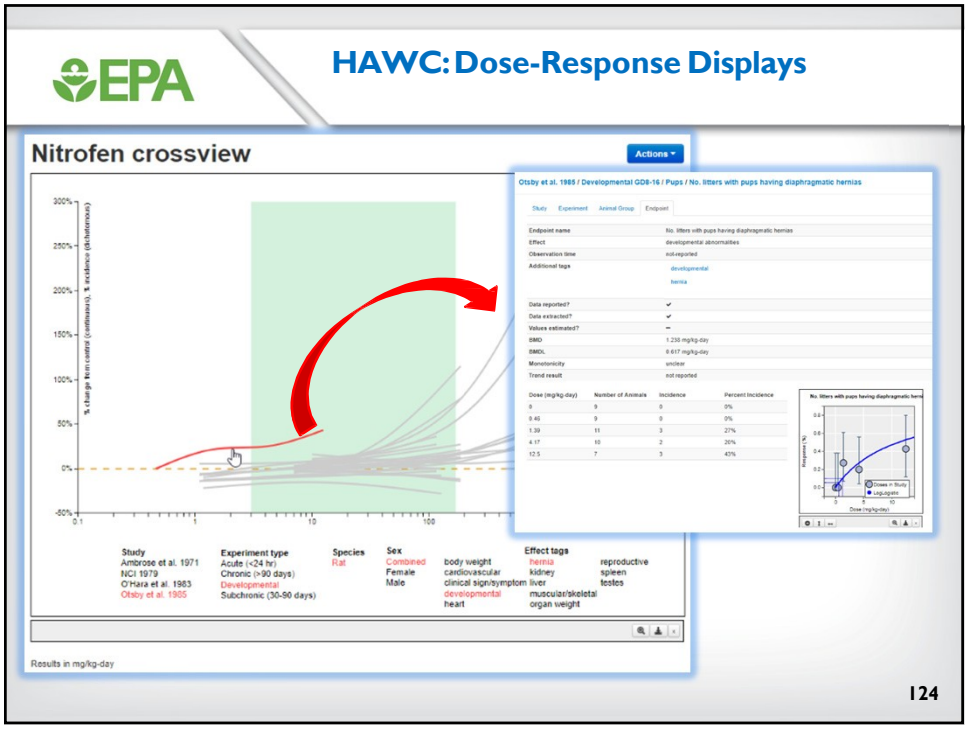


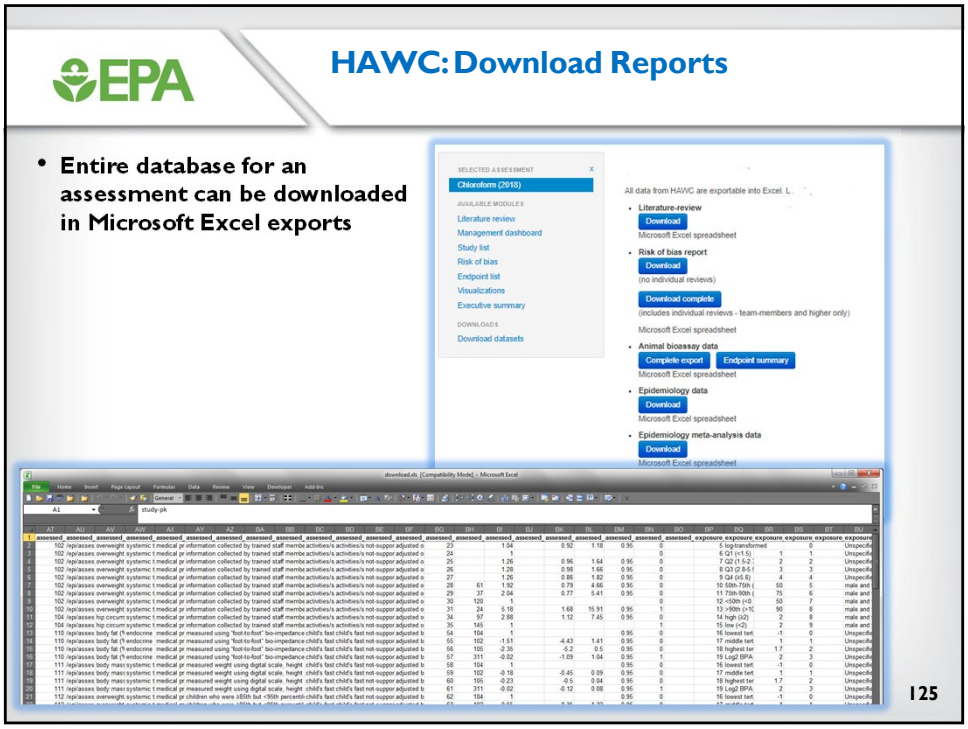






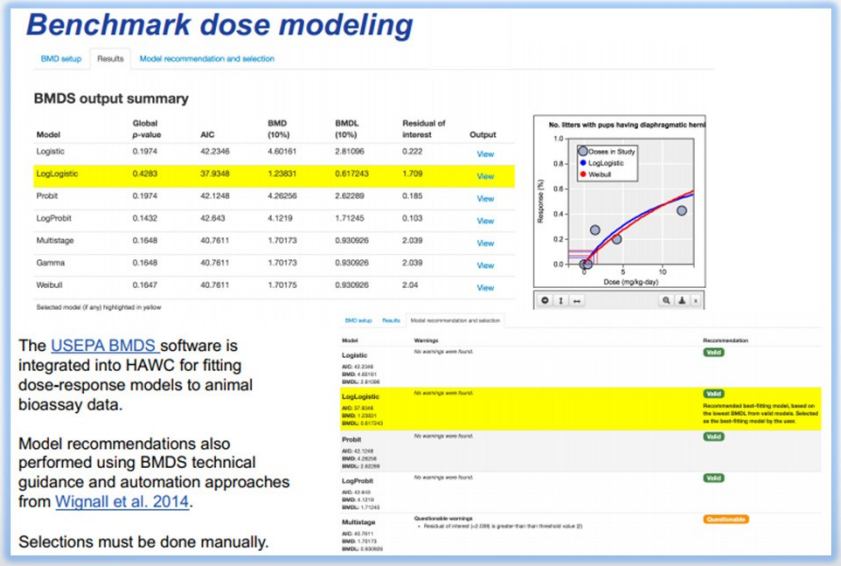
Appendix C





Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

HAWC Benchmark Dose Modeling



126

Advantages

- **Structured extraction to promote consistency and completeness**
- **Free, open source and customizable**
- **Enhance opportunities for database interpretability**
- **Integration with automated data-extraction tools**
- **Web-based to promote team collaboration**
- **Ability to export data files promotes further analysis of findings and quantification (in assessments or for methods development)**
- **Creates possibilities for web-based, interactive reports**

127

IRIS has Addressed the Major NAS 2014 Recommendations

NAS 2014 Topics	IRIS Process Improvements
Looking Forward	• Specialized software tools for efficiency and more user friendly and transparent formats for evidence display have been adopted • Strategic planning on use of text and data-mining tools and automation • Specialized tools facilitate transparent documentation, consistency across assessments, and database interoperability

See Demonstrations:

- SWIFT Review and SWIFT Active
- Health Assessment Workspace Collaborative
- Heath Effects Research Online

128

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

NAS 2014: Three High Priority (Box 8-1) Recommendations on Quantification

- **TOXICITY VALUES: "EPA should develop criteria for determining when evidence is sufficient to derive toxicity values."**
  - Overall hazard conclusions inform decision whether to develop toxicity values.
  - Better documenting considerations on which studies are carried forward to dose-response.
- **POINTS OF DEPARTURE (PODs): "EPA should clearly present two dose-response estimates: a central estimate (such as a maximum likelihood estimate or a posterior mean) and a lower-bound estimate for a POD from which a toxicity value is derived."**
  - Central estimates (MLEs) of BMDs provided in IRIS assessments along with BMDLs.
  - Will start to use WHO/IPCS approach to characterize distributions in final values.
  - Model averaging to characterize model uncertainty.
- **QUANTITATIVE CAPABILITIES: "EPA should expand its ability to perform quantitative modeling of evidence integration; in particular, it should develop the capacity to do Bayesian modeling of chemical hazards. ...The Committee emphasizes that... IRIS assessments should not be delayed while this capacity is being developed."**
  - Meta-analysis of human and animal studies increasing: hazard decisions and dose-response.
  - Bayesian methods are being explored to help characterize uncertainty.
  - New approach methods and assays are increasingly being evaluated quantitatively.

130

Evidence Integration Conclusions Inform when to Develop Toxicity Values

Evidence integration conclusion	Quantitative toxicity value provided?
Strongest conclusion for a human health effect (for cancer, a descriptor of *Known*)	Yes.
Moderately strong conclusion for a human health effect (for cancer, a descriptor of *Likely*)	Yes.
Weakest conclusion for a human health effect (for cancer, a descriptor of *Suggestive*)	Determined by situation (e.g., may provide values when useful for decision purpose and the evidence includes a well-conducted study)
Inadequate information	No, although bounding estimate from a study that does not show positive results can be derived where useful for decision purpose.
Strong support for no human health effect	No.

131

Decision-Making for Advancing Studies to Develop Toxicity Values

IRIS has further clarified the considerations that inform the selection of studies to estimate human dose-response relationships (next slide).

- IRIS continues to find that this decision process is not reducible to a formula.
- Expert judgment is essential for judging the relative merits of individual studies and which studies support more integrative quantitative analyses (e.g., meta-analysis).
- IRIS must often utilize studies with a range of attributes and levels of reporting. For example, the available studies on many mission-critical chemicals do not provide data on an individual subject basis.
- For full transparency, IRIS continues to emphasize documentation of the factors it weighed in emphasizing certain studies, or combinations of studies, over others.

132

More Explicitly Defining the Attributes IRIS Uses to Evaluate Studies for Derivation of Toxicity Values

In addition to qualitative study evaluation judgments (i.e., *medium* or *high* confidence studies are preferred), studies are assessed across several study attributes

Example Primary Considerations for Selection of Studies for Derivation of Toxicity Values			
Study attribute		**Human studies**	**Animal studies**
Test species		**Human data are generally preferred** to eliminate interspecies extrapolation uncertainties (e.g., in toxicodynamics and specific health outcomes).	**Animals that respond most like humans are preferred**. Outcomes associated with species known to show differences in sensitivity can provide support with suitable qualification.
Human relevance of the exposure paradigm	**Exposure route**	Studies involving **typical human environmental exposure routes are preferred** (e.g., oral, inhalation). A validated toxicokinetic model can be used to extrapolate across exposure routes.	
	Exposure duration	For chronic toxicity values, **chronic or subchronic studies are preferred**. **Exceptions exist** (e.g., when a population or lifestage is more sensitive during a particular time window)	
	Exposure levels	**Exposures near the range of typical environmental human exposures are preferred. Studies with a broad exposure range and multiple exposure levels are preferred** to the extent that they can provide information about the shape of the exposure-response relationship* and facilitate extrapolation to more relevant (generally lower) exposures.	
Susceptibility		Studies that yield **risk estimates in the most susceptible groups are preferred**. Inclusion of design features in the analysis (e.g., matching procedures, blocking; covariates or other procedures for statistical adjustment) **that adequately address the relevant sources of potential critical confounding for a given outcome are preferred**.	

*U.S. EPA Benchmark Dose Technical Guidance (2012)

133

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

IRIS Assessments Are Providing Central MLE Estimates of BMDs Along with BMDLs

Table 2-1. Summary of derivation of PODs

Endpoint and reference	Species/ sex	Model[a]	BMR	BMD mg/kg-d	BMDL mg/kg-d	POD mg/kg-d	POD$_{Adj}$[b] mg/kg-d	POD$_{HED}$[c] mg/kg-d
Developmental								
Neurobehavioral changes: Open field crossed squares at PND 69 Chen et al. (2012)	Male and Female Sprague-Dawley rats	Exponential 4	1 SD	0.23	0.11			
Neurobehavioral changes: Elevated plus maze open arm entries at PND 70 Chen et al. (2012)	Female Sprague-Dawley rats	Exponential 4	1 SD	0.21	0.092	0.092[d]	0.092	0.092
Neurobehavioral changes: Morris water maze hidden platform trial escape latency at PNDs 71–74 Chen et al. (2012)	Male and Female Sprague-Dawley rats	Hill CV Hill CV Hill CV Hill NCV	1 SD (9 sec)	PND71: 0.49 PND72: 0.33 PND73: 0.27 PND74: 0.23	0.16 0.16 0.12 0.13			
Cardiovascular effects at PND 53 Jules et al. (2012)	Long-Evans rats	LOAEL (0.6 mg/kg-d) (15% ↑ in systolic blood pressure; 33% ↑ in diastolic blood pressure)				0.6	0.6	0.15

Recent animal study example to the left: **Benzo[a]pyrene (EPA, 2017)**

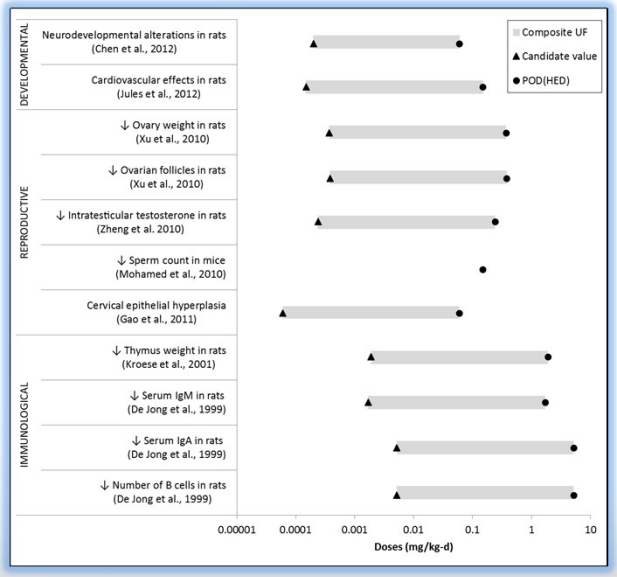
p. 2-8 Toxicological Review of Benzo[a]pyrene https://cfpub.epa.gov/ncea/iris/iris_documents/documents/toxreviews/0136tr.pdf

Recent epidemiology example: **Ethylene oxide (EPA, 2016)**

p. 4-109 Toxicological Review of the Inhalation Carcinogenicity of Ethylene Oxide. (EPA, 2016) https://cfpub.epa.gov/ncea/iris/iris_documents/documents/toxreviews/1025tr.pdf

134

IRIS is also Presenting Arrays of Candidate Toxicity Values



Benzo[a]pyrene (EPA, 2017)

135

Improvements in Characterizing Uncertainty

1) Model Averaging: characterizing model uncertainty

- Currently evaluating several methods
- Approach for dichotomous data expected to undergo peer review in 2018

$$\Pr(BMD \mid D) = \sum_{i=1}^{9} \pi_i \Pr(BMD \mid M_i, D)$$

Posterior Distribution of the BMD

$$\alpha = \int_{-\infty}^{BMD_\alpha} \Pr(BMD \mid D) dBMD$$

Calculation of the BMDL

136

Improvements in Characterizing Uncertainty

2) Distributions and Central Estimates: characterizing uncertainty in the human toxicity value

- WHO/IPCS guidance (IPCS, 2014)
- Risk-specific doses in terms of ranges, explicitly described:
  - Effect magnitudes
  - Confidence levels
  - Human population incidence rates.
- A probabilistic approach to adjustments from animal to human; a framework for refining toxicity values.

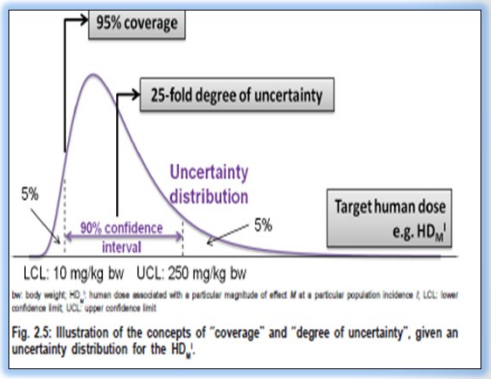


Fig. 2.5: Illustration of the concepts of "coverage" and "degree of uncertainty", given an uncertainty distribution for the HD_M^I.

137

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

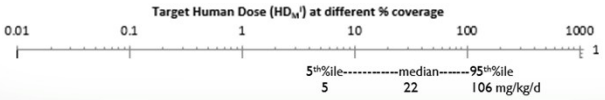
Improvements in Characterizing Uncertainty

WHO/IPCS Approach:

IRIS intends to provide such calculations along with traditional Reference Values:

- Confidence intervals on risk-specific doses
- Central estimates
- Estimates of incidence as a function of dose
- Use of appropriate probability math for uncertainty adjustments (instead of UFs) to allow for a more probabilistic and scientific value for use in risk assessment

By characterizing ranges of risk-specific doses, this provides more than a "conservative" estimate (it provides useful context by estimating the full distribution)



138

Use of Quantitative Modeling to Inform Evidence Integration

Meta-Analysis:

Increasingly Being Used to Interpret Sets of Results across Similar Populations

- Formal tools continue to be used to combine similar human epidemiology studies to improve decisions about hazard and about slope of dose-response.
- These approaches have also been used to better understand animal data that differ between studies of similar species and endpoints.
- As software tools and best practices become more common and easier to apply to environmental health studies, IRIS intends to consider their use more routinely.

Other examples: Libby Amphibole Asbestos (2014) and Trimethylbenzene analysis (Davis and Kraft, 2017) – see poster session; Arsenic assessment (in process)

139

Use of Quantitative Modeling to Inform Evidence Integration

Bayesian Approaches:

More Frequent Use Across Different Applications, and Research is Ongoing

- **Characterizing Uncertainty**
  - Bayesian approaches were used to characterize uncertainty in PBPK modeling and evaluate inter-related model inputs (Perchlorate peer review, 2018).
  - Bayesian Analysis is compatible with the WHO/IPCS Approach for characterizing uncertainty
- **Model Averaging**
  - Bayesian approaches are being applied to individual BMD models, and then model averaging is used to characterize uncertainty
- **Meta-Analysis**
  - Bayesian meta-analysis is currently being used to evaluate arsenic epidemiology studies
- **Bayesian Networks** (**exploratory research is currently underway**)
  - Possess the potential to integrate across evidence streams and bridge data gaps, borrowing strength from diverse data.
  - Software and mathematics are currently available.

140

Future work to better meet Agency needs for "benefits analysis"

Economics benefits analysis would ideally estimate incidence resulting from different decision options.

- We have provided human dose response functions from some analyses based on epidemiology data. (Evaluation of the Inhalation Carcinogenicity of Ethylene Oxide, EPA, 2016).

IRIS is also evaluating analogous predictions from animal data that could inform benefits analysis, including modifications of the IPCS approach.

141

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

Advancing Application of New Approach Methods (NAM) and Data in HHRA

- Over the past decade, several reports, books, resource documents, etc. have been published regarding the use of New Approach Methods (NAM) across the human health risk assessment paradigm (i.e., shifting the paradigm)
- Numerous labs, centers, workgroups, and initiatives across federal, private, and academic institutions have been formed to advance NAM

- EPA/ORD/NCEA, in conjunction with partners (e.g., NCCT, NTP) has been actively engaged in the conceptualization and evaluation of NAM across a broad landscape of HHRA applications

142

NAM Toolbox to Date

• **Data-mining**: ToxRefDB-comprehensive collection and collation of extant hazard and exposure data –(Martin et al. 2009. Env Health Perspect 117: 392-399)

• **Chemoinformatics**: structure-activity/read-across; QSAR –(Wang et al. 2012. Regul Toxicol Pharmacol 63: 10-19; Craig et al. 2014. J Appl Toxicol 34: 787-794)

• **High-Throughput (HT) Exposure modeling**: ExpoCast –(Egeghy et al. 2016. Env Health Perspect. 124(6):697-702)

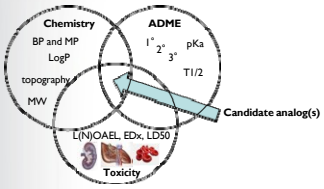
• **HT Toxicokinetics**: in vitro to in vivo (IVIVE) modeled dosimetry –(Wambaugh et al. 2015. Tox Sci 147: 55-67)

• **Bioactivity**: short-term animal; cell-free and/or cell-based HT assay data – (Judson et al. 2011. Chem Res Toxicol 24: 451-462; Dean et al. 2017. Tox Sci 157(1):85-99)

• **Adverse Outcome Pathway (AOP)**: expert-driven identification of signal transduction pathways along the exposure to outcome continuum. –(Edwards et al. 2016. J Pharmacol Exp Ther. 356(1):170-181)

143

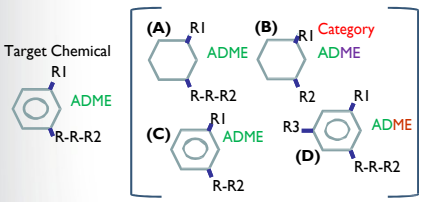
Chemoinformatics



Expert-driven Read-Across

Data-poor chemicals
- Inferred/interpolated hazard
- Surrogate based POD and subsequent derivation of RfVs

IRIS-type chemicals
- Data-gap filling
- Augment WOE
- Potential for reducing uncertainties



ADME = Absorption, Distribution, Metabolism, Elimination

Category approach

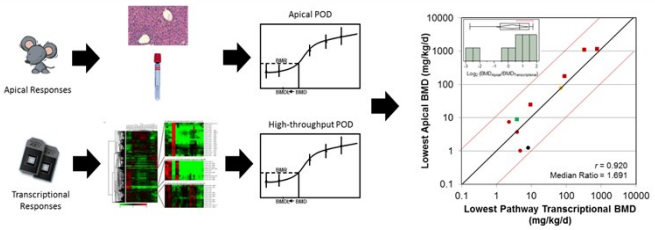
Data-poor chemicals
- Data-gap filling
- Extrapolated hazard
- Less applicable for quantitative assessment currently

IRIS-type chemicals
- Data-gap filling
- Augment WOE
- Foundational member of category (i.e., anchor chem)

- Similarity in structure and physicochemical properties between a chem of concern and a population of analogs
- Robustness of approach dependent on density of analogs populating a category
- Highly reliant on WOE supporting toxicity endpoints across category
- Presumes common Adverse Outcome Pathway or Mode of Action across category members

144

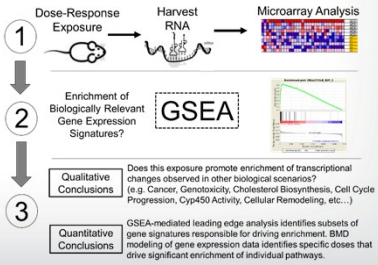
Bioactivity (e.g., transcriptomics)



- Close relationship between genotype/phenotype across two different routes of exposure, rodent species, and multiple target tissues
- In vitro?? Will need to optimize metabolism protocols; integrate IVIVE

Data-poor chemicals
- Evidence base for hazard
- Empirical dose-response based on pathway perturbations
- Reduce need for longer-term animal studies

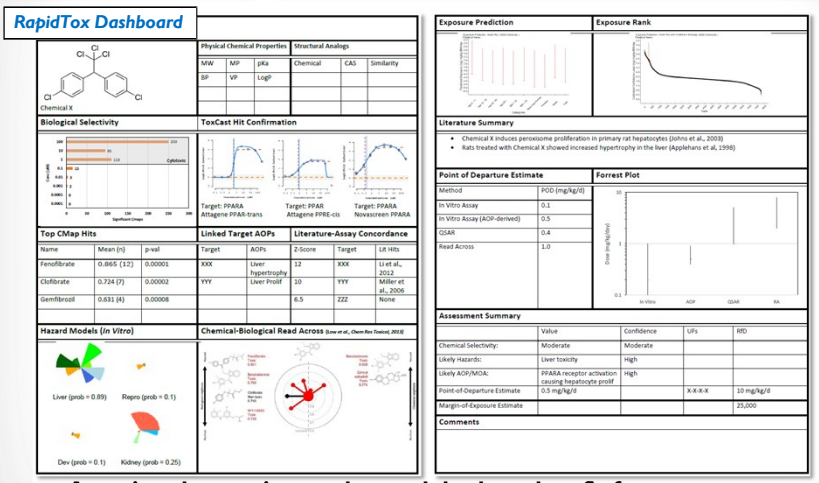
IRIS-type chemicals
- Augment WOE (e.g., MOA/AOP)
- Opportunity to alert off-target effects
- Potential for reducing uncertainties



145

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

Integrated Application to Risk Assessment



- **Associated narrative can be modular based on fit-for-purpose**
- **Systematic WOE always, but can be graded based on decision context**
- **Characterization of qualitative and quantitative uncertainties**

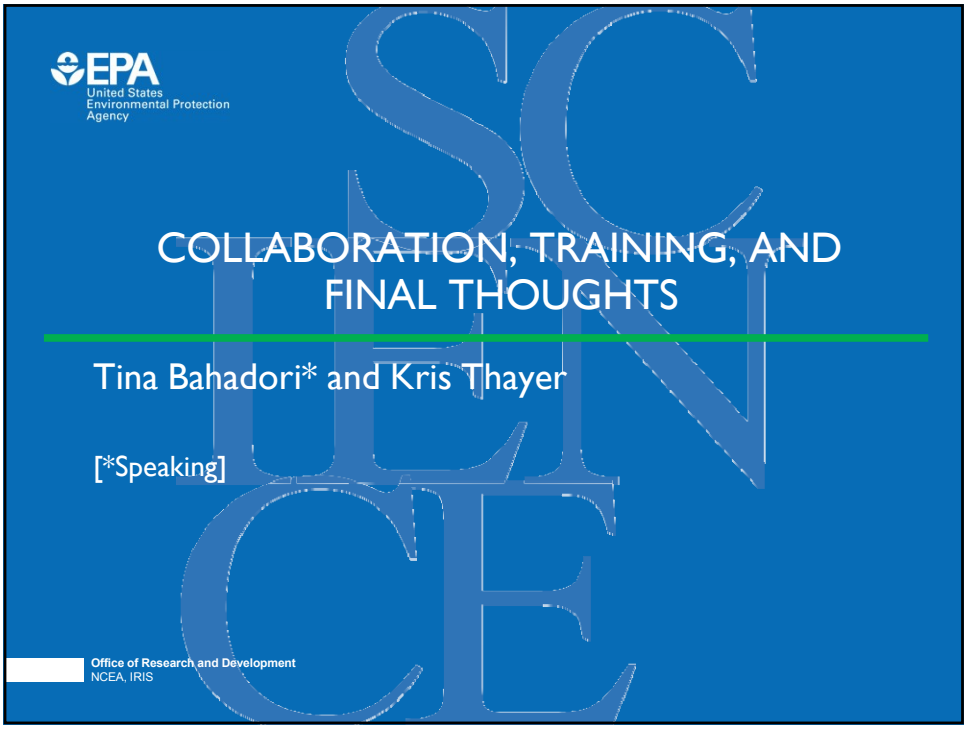
146

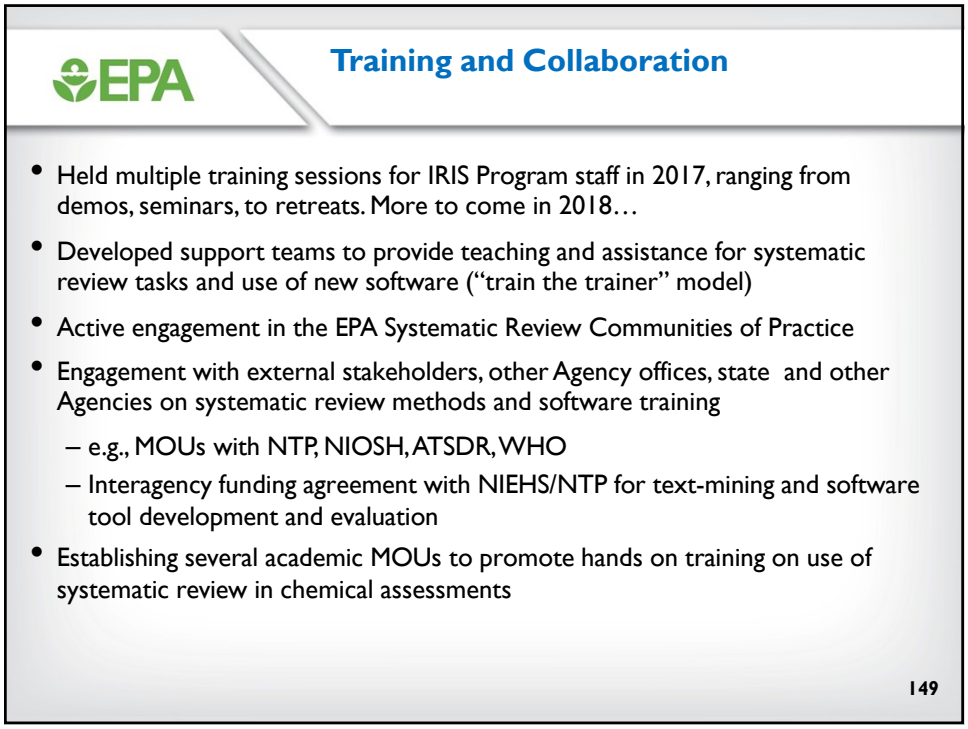
IRIS has Addressed the Major NAS 2014 Recommendations

NAS 2014 Topics	IRIS Process Improvements
Evidence Integration for Hazard Identification (Chapter 6) and Derivation of Toxicity Values (Chapter 7)	• Developing and applying quantitative tools in support of evidence synthesis and integration, including meta-analytical approaches • Expanded development and use of more advanced quantitative methods in software tools, such as BMDS • Developed more explicit criteria for deriving toxicity values, including the intent to derive quantitative toxicity values when IRIS reaches one of the stronger evidence integration conclusions, as well more specific criteria for the evaluation of individual studies • Providing MLE estimates of BMDs, along with BMDLs • Applying and exploring quantitative approaches to better characterize uncertainty, including probabilistic and Bayesian approaches
Future Directions (Chapter 8 "Lessons Learned" and "Looking Forward")	• Quantitative assessment methods will be updated in a continuing, strategic fashion, including capacity building (e.g., training; evolving best practices) for current approaches including meta-analysis, probabilistic analyses, and Bayesian methods

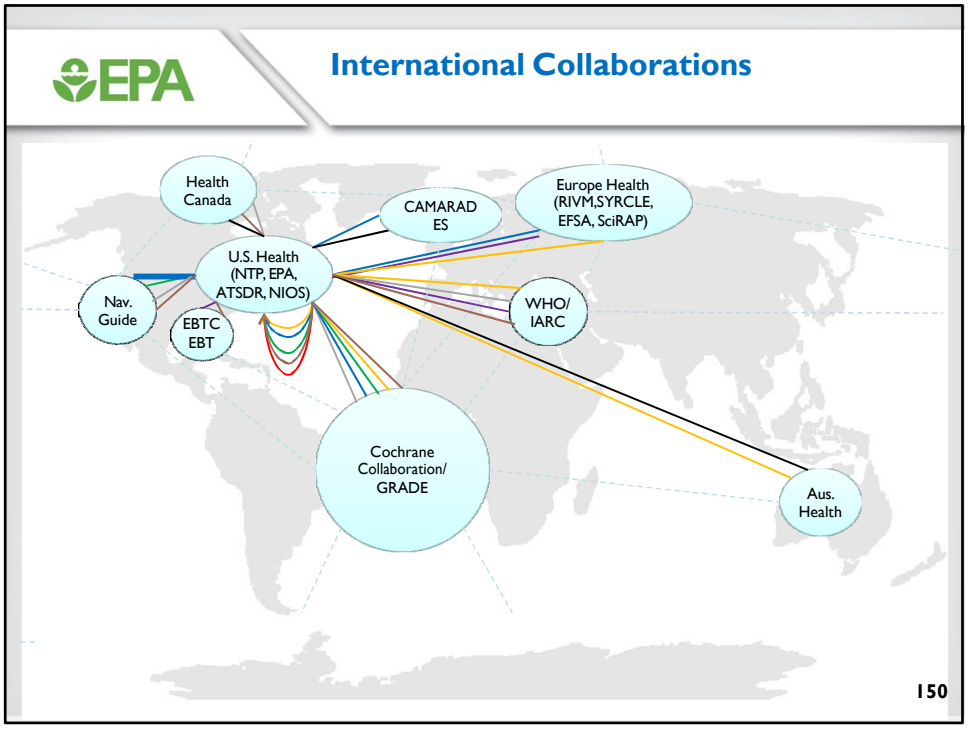
147

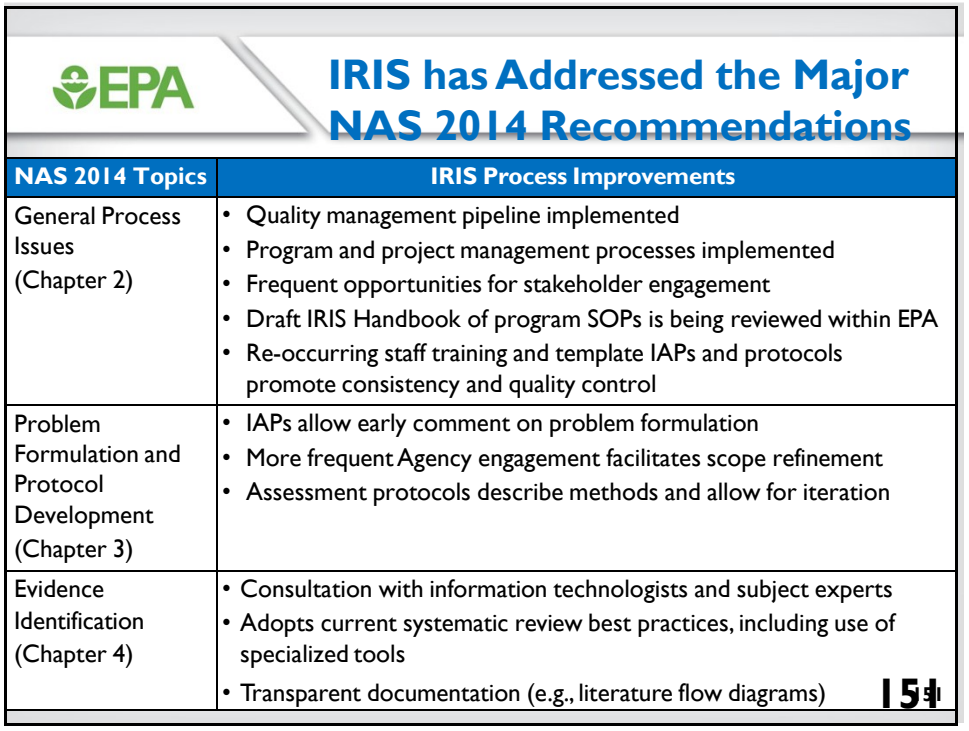
Appendix C





Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation





IRIS has Addressed the Major NAS 2014 Recommendations

NAS 2014 Topics	IRIS Process Improvements
Evidence Evaluation (Chapter 5)	• Individual studies are evaluated for reporting quality, risk of bias, and sensitivity • Decisions and supporting rationale are clearly documented • Study evaluations impact subsequent assessment decisions
Evidence Integration for Hazard Identification (Chapter 6)	• Structured frameworks provide transparency in expert judgments across human, animal, and mechanistic studies (based on Hill) • Standardized templates documenting key evidence integration decisions have been developed (evidence profile tables) • Developing and applying quantitative tools in support of evidence synthesis and integration, including meta-analytical approaches • Expanded development and use of more advanced quantitative methods in software tools, such as BMDS

152

IRIS has Addressed the Major NAS 2014 Recommendations

NAS 2014 Topics	IRIS Process Improvements
Derivation of Toxicity Values (Chapter 7)	• Developed more explicit criteria for deriving toxicity values, including the intent to derive quantitative toxicity values when IRIS reaches one of the stronger evidence integration conclusions, as well more specific criteria for the evaluation of individual studies • Providing MLE estimates of BMDs, along with BMDLs • Applying and exploring quantitative approaches to better characterize uncertainty, including probabilistic and Bayesian approaches

153

Progress Toward Transforming the Integrated Risk Information System Program: A 2018 Evaluation

IRIS has Addressed the Major NAS 2014 Recommendations

NAS 2014 Topics	IRIS Process Improvements
Future Directions (Chapter 8 "Lessons Learned" and "Looking Forward")	• Processes being implemented include flexibility to incorporate evolving methods in systematic review and risk assessment • Increased collaboration with federal partners and international experts prevents duplication of effort and maintains cutting edge approaches • Current research efforts and training serve to ensure that methods and staff are able to adapt to changing scientific contexts and sources of evidence, including new and emerging data types • Specialized software tools for efficiency and more user friendly and transparent formats for evidence display have been adopted • Strategic planning on use of text and data-mining tools and automation • Specialized tools facilitate transparent documentation, consistency across assessments, and database interoperability • Quantitative assessment methods will be updated in a continuing, strategic fashion, including capacity building (e.g., training; evolving best practices) for current approaches including meta-analysis, probabilistic analyses, and Bayesian methods

154

Appendix D

Posters by US Environmental Protection Agency

- D-1: New Approach Methods in Human Health Risk Assessment
- D-2: Combining Data within Species: Meta-analysis in IRIS
- D-3: Systematic Evaluations of Physiologically-Based Pharmacokinetic Models for Human Health Risk Assessment
- D-4: Male Reproductive Toxicity in Animal Studies of Diisobutyl Phthalate (DIBO): A Case Study Application of Systematic Review Approaches
- D-5: Male Reproductive Toxicity in Epidemiology Studies of Phthalates: A Case Study Application of Systematic Review Approaches
- D-6: Quantitative Evaluation of Uncertainty: APROBA and Beyond
- D-7: Mode of Action and Human Relevance Evaluation of Dibutyl Phthalate (DBP)-Induced Male Reproductive System Toxicity
- D-8: EPA Dose-Response & Related Software – New & Future Developments
- D-9: Evidence Profile Table for DIBP and Male Reproductive Toxicity
- D-10: A New Bayesian Approach to Combining Different Species Data

New Approach Methods in Human Health Risk Assessment

Jason C. Lambert

U.S EPA, Office of Research and Development, National Center for Environmental Assessment - Cincinnati

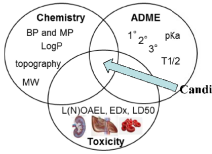
Integration of New Approach Methods-Theory

- Chemicals nominated for Human Health Risk Assessment (HHRA) have widely varying hazard and dose-response databases
- Integration of New Approach Methods (NAM) is therefore fit-for-purpose along a decision-based gradient:
  - Data-poor chemicals⇨ NAM may be a driver
  - Data-rich chemicals ⇨ NAM fills data gaps
  - Same/similar assays, same/similar data can be used in different ways to answer specific questions
- NAMs currently being integrated or evaluated in EPA HHRA contexts include:
  - Read across (expert-driven; category-based)
  - Transcriptomics (*in vivo* short-term animal)
  - High-throughput bioactivity
- Although not NAM per se, transparency principles of systematic review and integration of toxicity pathway [e.g., AOP or MOA) information also paramount

Read-Across

Expert-driven read-across

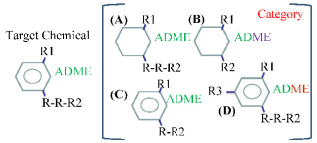
- 'Many-to-one approach'
- Approach is based on evidence across three information tiers (e.g., structural and physicochemical; toxicokinetic and toxicity/bioactivity) to select analog(s)
- Hazard and dose-response information (e.g., point-of-departure [POD]) from single best analog used as surrogate for target chemical



Candidate analog(s)⇨ POD from single best analog adopted as surrogate for data-poor target chemical

Category based read-across

- 'One-to-many' approach
- Based primarily on structural and physicochemical properties
- Robustness of approach dependent on density of analogs populating a category
- Highly reliant on weight-of-evidence supporting toxicity endpoints across category
- Presumes common AOP or MOA across category members



Tox for category:
(A)= liver
(B)= kidney, liver
(C)= liver
(D)= GI, liver, kidney

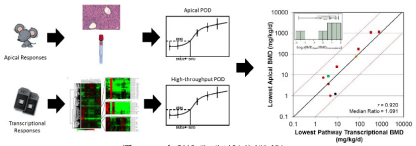
ADME: = Absorption, Distribution, Metabolism, Elimination

HHRA application(s)

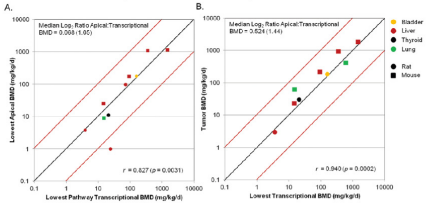
- Provisional Peer-Reviewed Toxicity Value (PPRTV) assessments; Superfund Technical Support memos to EPA Regions

Transcriptomics

Transcriptional perturbations and apical endpoints for both cancer and noncancer are evaluated in same organ tissues following short-term (e.g., 2-week) exposures

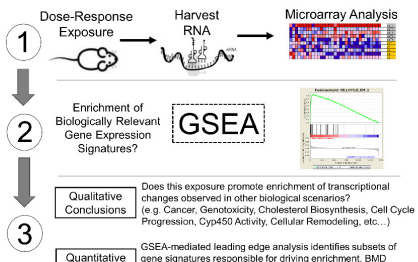


(Thomas et al., 2013, *Tox Sci* 134(1):180-194)



- Transcriptional pathway-based points-of-departure (PODs) from short-term *in vivo* assays were within 2-3 fold of both non-cancer (A) and cancer (B) apical PODs across different species, routes of exposure, durations of exposure, and target organ tissues
- Major challenge: relevance of transcriptional pathway perturbations to target organ toxicity?

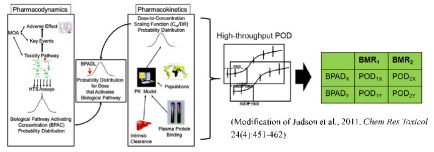
GSEA: Identifying Biologically-Relevant Transcriptional Alterations



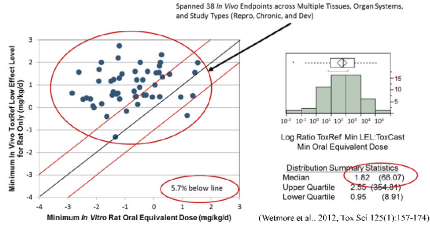
(Image courtesy of Dr. Jeffry Dean, EPA/ORD/NCEA-Cincinnati)

High-throughput Bioactivity

Integration of in vitro biological activity data (e.g., ToxCast/Tox21) and reverse toxicokinetic in vitro to in vivo extrapolation may facilitate identification of oral equivalent doses that can be benchmark dose modeled for identification of HTP-based PODs



(Modification of Judson et al., 2011, *Chem Res Toxicol* 24(4):451-462)



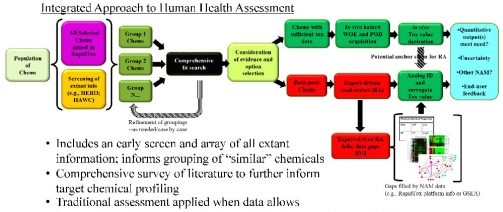
(Wetmore et al., 2012, Tox Sci 125(1):157-174)

HHRA application(s)

- Superfund Technical Support memos to EPA Regions; bioactivity information used as qualitative support for augmenting weight-of-evidence in analog(s) selection

Bringing it all together

Integrated Approach to Human Health Assessment



- Includes an early screen and array of all extant information; informs grouping of "similar" chemicals
- Comprehensive survey of literature to further inform target chemical profiling
- Traditional assessment applied when data allows
- Expert-driven read-across when hazard/dose-response data are lacking
- Integration of information from NAM data streams to fill gaps

**The collective Agency efforts presented here are in response to the NAS' suggestion to put research/processes in place to adapt to new and emerging methods

U.S. Environmental Protection Agency
Office of Research and Development

The views expressed in this poster are those of the author and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency

EPA United States Environmental Protection Agency

Combining data within species: Meta-analysis in IRIS

J. Allen Davis[1] and Leonid Kopylev[2]

[1] US EPA, Office of Research and Development, National Center for Environmental Assessment - Cincinnati
[2] US EPA, Office of Research and Development, National Center for Environmental Assessment - Washington

Introduction

Often, human health risk assessments have relied on qualitative approaches for hazard identification, which involves weight of evidence determinations that integrate evidence across multiple studies. In 2014, the National Research Council recommended that IRIS develop and apply quantitative approaches for evidence integration, including the application of meta-analyses to animal and human data, to help summarize and evaluate the results of a systematic review. In the meta-analytic approach, a pooled effect size is calculated after consideration of multiple potential confounding factors in order to determine whether the entire database under consideration indicates a chemical is a hazard. Two examples demonstrate approaches used in IRIS assessments: TMB (trimethylbenzene) neurotoxic hazard and pleural plaques effect on lung function.

Trimethylbenzene and pain sensitivity: methods

- A publically available, comprehensive literature search was performed in support of the IRIS Toxicological Review of trimethylbenzenes (TMBs)
- Six neurotoxicity studies were found that investigated decreased pain sensitivity following exposure in individual TMB isomers or a mixture thereof (i.e., C-9 fraction) - studies differed in testing time, test agent, and application of foot shock
- Qualitative hazard identification concluded the pain sensitivity was a hazard and that testing time mainly influenced observation of effect
- Methods outlined in Vesterinen et al. (2014) and Viechtbauer (2010) were applied using the Metafor R package
  - Random and mixed-effects models were run
  - Effect sizes were calculated as standardized mean differences
  - Hedge's G was used to account for bias due to small sample sizes
  - Restricted maximum likelihood was used to calculate total heterogeneity to prevent underestimated/biased estimates of variance
  - Publication bias, normality of residuals and sensitivity analyses were investigated

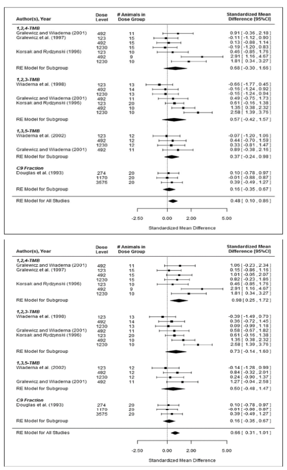


Figure 1. Forest plots for pain sensitivity studies. (A) Pre-foot shock; (B) post-foot shock

Trimethylbenzene and pain sensitivity: results

Table 1– Meta-regression results for TMB and mixture studies included in analysis

Moderating Variable	Effect Size (95% CI) Model 1	Model 2	Number of Animals	Number of Groups	Adjusted R^2	Q_E p-value[a]	Q_M p-value[b]
Pre-Foot Shock Results							
Pooled Estimate	0.479 (0.103-0.855)		283	21			
Lab							
Nofer	0.550 (0.129-0.971)*	Ref	223	18	0.0%	0.006	0.423
Douglas	0.159 (-0.751-1.069)	-0.392 (-1.394-0.610)	60	3			
Testing Time							
0 days	1.331 (0.813-1.849)***	Ref	69	6	93.9%	0.333	0.002
1 day	0.158 (-0.424-0.740)	-1.173 (-1.95.2- -0.394)**	60	3			
50 days	0.106 (-0.258-0.470)	-1.225 (-1.858- -0.591)***	154	12			
Isomer							
1,2,4-TMB	0.647 (-0.099-1.393)	Ref	85	7	0.0%	0.003	0.845
1,2,3-TMB	0.573 (-0.121-1.266)	-0.074 (-1.093-0.944)	91	7			
1,3,5-TMB	0.386 (-0.575-1.346)	-0.261 (-1.477-0.955)	47	4			
C9 fraction	0.159 (-0.828-1.145)	-0.488 (-1.725-0.749	60	3			
Dose	0.0001 (-0.0004-0.0006) for each 10 mg/m³ increase				0.0%	0.004	0.718
Post-foot Shock Results							
Pooled Estimate	0.663 (0.314-1.012)		283	21			
Lab							
Nofer	0.772 (0.397-1.146)***	Ref	223	18	22.02%	0.041	0.153
Douglas	0.159 (-0.619-0.936)	-0.613 (-1.476-0.250)	60	3			
Foot-shock							
Yes	0.911 (0.395-1.426)**	Ref			0.00%	0.024	0.196
No	0.467 (0.004-0.929)*	-0.444 (-1.137-0.249)					
Testing Time							
0 days	1.361 (0.812-1.909)***	Ref	69	6	72.65%	0.238	0.013
1 day	0.158 (-0.467-0.784)	-1.203 (-2.034- -0.371)**	60	3			
51 days	0.459 (0.074-0.844)*	-0.902 (-1.572- -0.232)*	154	12			
Isomer							
1,2,4-TMB	1.004 (0.346-1.662)**	Ref	85	7	0.0%	0.030	0.410
1,2,3-TMB	0.722 (0.128-1.316)*	-0.282 (-1.168-0.604)	91	7			
1,3,5-TMB	0.5.5 (-0.330-1.361)	-0.489 (-1.560-0.582)	47	4			
C9 fraction	0.159 (-0.665-0.982)	-0.846 (-1.900-0.209)	60	3			
Dose	0.0001 (-0.0004-0.0005) for each 10 mg/m³ increase				0.0%		0.807

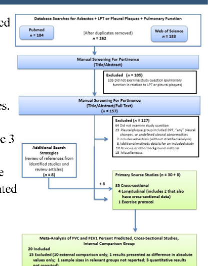
Model 1 – Intercept suppressed, effects estimated separately for each variable level, Model 2 – estimated effects are relative to the variable level identified as the Reference.
Individual categorical moderators are statistically significant at the following levels: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$
[a] Q_E is the test for residual heterogeneity; p values < 0.05 possibly indicate that other moderators not included in the model may influence the measured outcome, specific to Model 2
[b] Q_M is the omnibus test of moderators; p values < 0.05 indicate that the moderators included in the model have a significant influence on the measured outcome, specific to Model 2

- Quantitative meta-analyses and meta-regressions supported original qualitative hazard identification determination – ***decreased pain sensitivity is a hazard in humans following exposure to trimethylbenzene isomers***
- Time of testing appeared to be the study-level variable that most strongly affected differing study results and explained the majority of inter-study heterogeneity

Pleural plaques effect on lung function: methods

A literature search was conducted using the PubMed and Web of Science databases with no publication date limitations. Studies were excluded if

- the plaques group included individuals with diffuse pleural thickening (DPT)
- undefined pleural or parenchymal abnormalities.



Each paper was reviewed independently by 2 of the 3 reviewers. In cases of disagreements, the 3rd reviewer reviewed the paper and participated in the consensus- building discussions. Reviewers evaluated potential limitations in 5 aspects of study design:

- selection of participants
- protocols for x ray or HRCT readings
- protocols for spirometry measurements;
- analytic approach
- considerations of smoking.

The Metaphor R package was used for the meta-analyses

- A random effects model was used for both FVC and FEV1
- To assess possible publication bias, funnel plots were evaluated. Additional sensitivity analyses were conducted to evaluate the potential effect of identified limitations on the meta-analyses results.

Pleural plaques effects on lung function: results

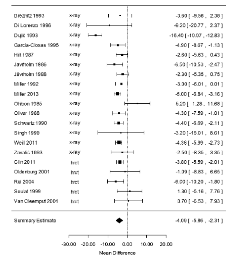


Figure 1 (FVC)

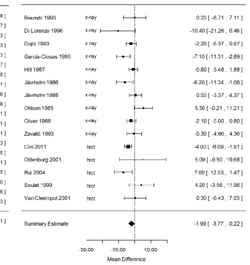


Figure 2 (FEV1)

- The summary effect estimates for both FVC and FEV1 are statistically significant, showing a change of −4.09 %pred (95% CI: −5.86, −2.31) and −1.99 %pred (95% CI: 3.77, 0.22), respectively (See Fig. 1 and Fig. 2)
- The results of larger studies are very consistent in showing a decrease in FVC (see Fig. 1). In contrast, fewer large studies are available for FEV1, and there is less consistency in the results (see Fig. 2).
- At the individual level, the decrement in FVC or FEV1 may or may not be noticeable for a given patient; while many with pleural plaques could have well-preserved lung function, there are some at the lower end of 'normal' lung function, for whom even a small additional decrement would result in an increased in disease severity (e.g., mild to moderate disease).
- At the population level, even small changes in the average of a distribution of lung function can result in a proportion of the exposed population shifted down into the lower "tail" of the distribution, into clinically significant lung function deficit region

Discussion

- Both human and animal data are amenable to quantitative synthesis via meta-analysis
- Studies need not be exactly the same, as long as results are reported in a consistent way or can be converted into a comparable format (e.g., use of standardized mean difference as effect metric)
- Use of free R software allows conducting meta-analysis
- Use of meta-analytic methods for hazard identification are in line with National Research Council (2014) recommendations for the development of quantitative hazard identification and evidence integration methods
- Applying meta-analysis and meta-regression methods will help to improve future risk assessments and ensure the use of the best available science

References:

Davis JA, Krall A. Quantitative meta-analytic approaches for the systematic synthesis of data and hazard identification: a case study of decreased pain sensitivity due to trimethylbenzene exposure. 2017. *Environmental Research.* 158: 598-609.

Kopylev L, Christensen KY, Brown JS, Cooper GS. A systematic review of the association between pleural plaques and changes in lung function. 2015. *Occupational and Environmental Medicine.* 72(8): 606-14.

Systematic Evaluations of Physiologically-Based Pharmacokinetic Models for Human Health Risk Assessment

Alan F. Sasso[1], Paul M. Schlosser[1]
[1]U.S. EPA, Office of Research and Development, National Center for Environmental Assessment

Background

- Physiologically-based pharmacokinetic (PBPK) models are tools for estimating absorption, distribution, metabolism, and elimination (ADME) of chemicals in the body
  - Quantify internal (tissue/organ) dose vs exposure
  - Facilitate dose-response analysis/human extrapolation
  - Use chemical- and species-specific data (unlike default $BW^{3/4}$ allometric scaling)
- Multiple alternative models or analyses in literature
  - "Being published is not enough": EPA thoroughly evaluates models based on scientific and technical criteria prior to use in an assessment
  - IRIS uses a structures approach to evaluate quality and usability
- The evaluation process stresses: (1) clarity in the documentation of model purpose, structure, and biological characterization; (2) validation of mathematical descriptions, parameter values, and computer implementation; and (3) evaluation of each plausible dose metric.
- **NAS (2014) recommendations addressed**
  - **Develop and expand use of formal quantitative methods in data integration for dose-response assessment and derivation of toxicity values**
  - **Develop tools for assessing risk of bias in different types of studies**

Identification and Inventory of PBPK Models

- A thorough literature search is conducted to identify existing PBPK models
- A summary report is prepared of available models and their possible utility for use **(scoping)**
  - This work is conducted by the **Pharmacokinetics Workgroup (PKWG)***, in conjunction with information specialists
- Table 1 outlines typical summary information presented for each model at the scoping phase

Table 1. Example animal PBPK inventory table for model scoping

Author	Smith et al. (2003)					
Contact Email	xxxxx@email.com					
Contact Phone	xxx-xxx-xxxx					
Sponsor	N/A					
Model Summary						
Species	Rat					
Strain	F344					
Sex	Male and female					
Life-Stage	Adult					
Exposure Routes	Inhalation	Oral	I.V.	Skin		
Tissue Dosimetry	Blood	Liver	Kidney	Urine	Lung	
Model Evaluation						
Language	ACSL 11.8					
Code Available	YES	Effort to Recreate Model		COMPLETE		
Code Received	YES	Effort to Migrate Code		SIGNIFICANT		
Structure Evaluated	YES					
Math Evaluated	YES					
Code Evaluated	YES. Issue (minor): Incorrect units listed in comments for liver metabolism (line 233). Issue (major): Mass balance error in stomach compartment.					
Available PK Data	Urine (cumulative amount excreted) and blood (concentration) time course data for oral (gavage) and inhalation (6hr/day for 4 days) exposure. In vitro skin permeation.					

***The Pharmacokinetics Workgroup (PKWG)** is convened by the National Center for Environment Assessment (NCEA) to support and promote consistent application of the best science in the field of mathematical modeling of pharmacokinetic processes and the data supporting it as applied in human health risk assessment. It is composed of scientists with specific expertise in the range of disciplines involved in the construction and development of pharmacokinetic models, evaluation of data supporting such models, statistical analysis of data and modeling results, and characterization of related uncertainty and variability.

Evaluation of PBPK Models

PBPK Model Scoping: Criteria A

- An evaluation of a model is required before accepting it for use in an assessment
  - Many models contain errors with varying degrees of impact on model predictions
- Initial judgments on the suitability of a model are separated into two categories: scientific and technical (Table 2)

Table 2. Evaluation criteria for PBPK models

Criteria	Example information
Scientific	Biological basis for the model is accurate. - Consistent with mechanisms that significantly impact dosimetry. - Predicts dose metrics expected to be relevant. - Applicable for relevant route(s) of exposure.
	Consideration of model fidelity to the biological system strengthens the scientific basis of the assessment relative to standard exposure-based extrapolation (default) approaches. - Can the model describe critical behavior, such as nonlinear kinetics in a relevant dose range, better than the default (i.e., $BW^{3/4}$ scaling)? - Is the available metric a better predictor of risk than default? Specifically, model-based metrics may correlate better than the applied doses with animal/human dose-response data. Degree of certainty in model predictions vs. default is also a factor. For example, while target tissue metrics are generally considered better than blood concentration metrics, lack of data to validate tissue predictions when blood data are available may lead to a choice of the latter.
	Principle of parsimony - Model complexity or biological scale, including number and parameterization of (sub)compartments (e.g., tissue or subcellular levels) should be commensurate with data available to identify parameters.
	Model describes existing PK data reasonably well, both in "shape" (matches curvature, inflection points, peak concentration time, etc.) and quantitatively (e.g., within a factor of 2–3).
	Model equations are consistent with biochemical understanding and biological plausibility.
Initial technical	Well-documented model code is readily available to EPA and public.
	Set of published parameters clearly identified, including origin/derivation.
	Parameters do not vary unpredictably with dose (e.g., any dose dependence in absorption constants is predictable across the dose ranges relevant for animal and human modeling).
	Sensitivity and uncertainty analysis has been conducted for relevant exposure levels (local sensitivity analysis is sufficient, though global provides more information). - If a sensitivity analysis was not conducted, the PKWG would suggest this as additional work before using the model in the risk assessment. - A sound explanation should be provided when sensitivity of the dose metric to model parameters differs from what is reasonably expected based on experience.

$BW^{3/4}$= body-weight scaling to the 3/4 power; PK = pharmacokinetic; PKWG = Pharmacokinetic Working Group

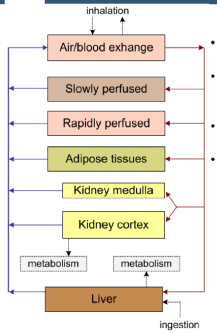
In-Depth Technical Evaluation: Criteria B

- Primarily address computational implementation and technical issues
- Only conducted on models that pass review for Criteria A
  - Criteria B evaluation is not possible without model code
- Model equations and parameters in computer codes match those in the manuscript or report
- Published figures/tables of model simulations are reproducible using the available code (within 10% of the publication).
- If errors in model code or parameters are found and corrected, the revised model must still be in agreement with data. Errors must be small enough to not invalidate the model, parameters, or assumptions.
  - Model predictions outside the range of the data are allowed to change by more than 10% of the original model or publication, since this would be considered a model correction.

Resource Considerations for PBPK Model Revision or Development: Criteria C

If existing models fail Criteria A or B, the potential value in implementing a PBPK in a risk assessment must be weighed against the time, effort, and possible expenses required to address model shortcomings.

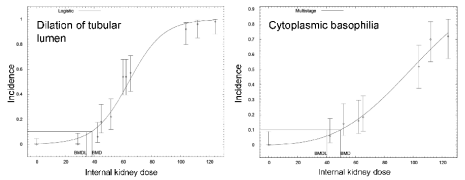
PBPK Evaluation Example: Chloroform



- Chloroform is a trihalomethane present in drinking water as a by product of disinfection
- Kidney and liver are target organs
  - Effects induced via production of reactive metabolites
- A PBPK model (left) was obtained during scoping and satisfied Criteria A
- Issues identified during in-depth technical evaluation:
  - Metabolic parameter derivation for the kidney cortex contained a units error, and the calculation was not performed consistently for humans and rodents
  - The volume ratio for kidney cortex and medulla was reversed in the code, and did not match the reported value or original reference

Upon evaluation under Criteria C, it was determined:

- Time and effort to correct the model was minimal
- Corrections led to little or no changes in model predictions of data
  - Estimates of the internal dose metric (kidney metabolism) changed significantly. Since there are no *in vivo* data available for this measure, this was considered a correction to the original model.
- Model was successfully revised by EPA, published as journal article (Sasso et al., 2013)



The revised PBPK model allows for improved quantitative dose-response modeling and data integration. Kidney endpoints can be evaluated across different routes of exposure and different species (Nagano et al., 2006, and Yamamoto et al., 2002). The figures above illustrate dose-responses for rats from multiple exposure routes (inhalation, oral, and combined inhalation+oral) on basis of PBPK-derived kidney dose.

Selected references

Mel.anahan et al. (2012). Physiologically based pharmacokinetic model use in risk assessment--Why being published is not enough. *Tox. Sci.*, 126: 5-15.

Nagano, et al. (2006). Enhancement of renal carcinogenicity by combined inhalation and oral exposures to chloroform in male rats. *J. Toxicol. Environ. Health Part A* 69, 1827–1842

Sasso et al. (2013). Application of an updated physiologically based pharmacokinetic model for chloroform to evaluate CYP2E1-mediated renal toxicity in rats and mice. *Tox. Sci.*, 131: 360-374.

Yamamoto, et al. (2002). Carcinogenicity and chronic toxicity in rats and mice exposed to chloroform by inhalation. *J. Occup. Health* 44, 283–293.

U.S. Environmental Protection Agency
Office of Research and Development

Male reproductive toxicity in animal studies of diisobutyl phthalate (DIBP): a case study application of systematic review approaches

Erin E. Yost[1], Susan Y. Euling[2], James A. Weaver[1], Brandiese E. J. Beverly[1], Nagalakshmi Keshava[1], Anuradha Mudipalli[1], Xabier Arzuaga[2], Todd Blessinger[2], Laura Dishaw[1], Andrew Hotchkiss[1], Susan L. Makris[2]

[1] US EPA National Center for Environmental Assessment, Research Triangle Park, NC; [2] US EPA National Center for Environmental Assessment, Washington, DC

Introduction

Diisobutyl phthalate (DIBP) is used as a plasticizer in a variety of industrial and consumer products. Although DIBP has been less widely studied compared to other phthalates, there is evidence that DIBP and its primary metabolite, monoisobutyl phthalate (MIBP), cause male reproductive toxicity. A recent systematic review of endocrine-related low-dose toxicity by the National Academies of Sciences (NAS) evaluated the effects of DIBP on three anti-androgenic outcomes [testosterone, anogenital distance (AGD), and hypospadias], and concluded that DIBP is a presumed human hazard based on decreased fetal testosterone in rodents exposed during gestation. The Integrated Risk Information System (IRIS) performed a systematic review of male reproductive effects of DIBP exposure that considered all outcomes and all life stages of exposure, following recommendations in the 2014 NAS review of the IRIS program. Here, we use studies that evaluated testosterone in male rodents exposed to DIBP or MIBP as a case study of the IRIS systematic review process. We also summarize the overall conclusions for male reproductive effects identified in the IRIS systematic review of DIBP, and compare these results to the findings of NAS.

Methods

Animal studies for DIBP or MIBP were identified by searching four online databases (PubMed, Web of Science, Toxline, and TSCATS2), using search terms designed to capture all potentially pertinent studies. The last update was in July 2017. Title/abstract screening was used to identify primary health effect studies that exposed non-human mammalian animals to any administered dose of DIBP or MIBP via oral, dermal, or inhalation routes. These studies were evaluated by at least two reviewers using the approach in Figure 1.

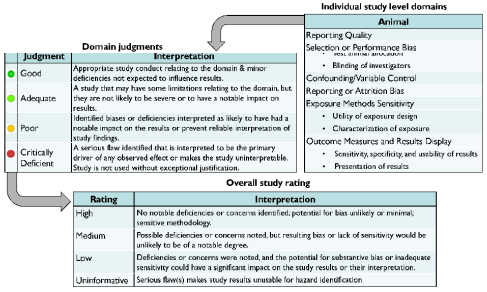


Figure 1. Study evaluation process

After study evaluation, the evidence for each health effect outcome was synthesized according to the developmental stage of exposure. Based on this synthesis, the evidence was assigned a conclusion of *robust, moderate, slight, indeterminate,* or *compelling evidence of no effect.* The ratings for individual outcomes were summarized into an overall conclusion for male reproductive effects using a structured framework (see Poster by Yost et al.).

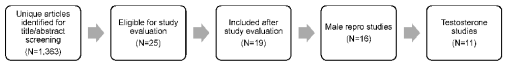


Figure 2. Abbreviated literature flow diagram

U.S. Environmental Protection Agency
Office of Research and Development

Results

Table 1. Animal studies of testosterone and DIBP or MIBP exposure. Of the 11 studies that evaluated testosterone in male rats or mice, 7 exposed animals during gestation and/or until weaning, and 4 were postnatal exposures of males near the time of puberty. The postnatal exposure studies had higher risk of bias because of reporting limitations, including uncertainty about the pubertal status of the test animals at the time of exposure.

Reference	Study description			Study evaluation										
	Population	Exposure	Outcome	Reporting quality	Test animal allocation	Blinding of investigators	Confounding / variable control	Reporting or attrition bias	Characterization of exposure	Utility of exposure design	Sensitivity, specificity, and usability of results	Presentation of results	Overall confidence
Borch et al. 2006	Rat (Wistar)	Diet GD 7-19	Fetal T prod/conc	G	G	A	G	G	A	A	G	G	High
Howdeshell et al. 2008	Rat (Sprague-Dawley)	Gavage GD 8-18	Fetal T prod	G	A	A	G	G	A	G	G	G	High
Saillenfait et al. 2017	Rat (Sprague-Dawley)	Gavage GD 13-19	Fetal T prod	G	G	A	G	G	G	G	G	G	High
Furr et al. 2014	Rat (Sprague-Dawley)	Gavage GD 14-18	Fetal T prod	G	G	A	G	A	A	G	G	G	High
Hannas et al. 2012	Rat (Sprague-Dawley)	Gavage GD 14-18	Fetal T prod	G	A	A	G	G	G	G	G	A	High
Hannas et al. 2011	Rat (Sprague-Dawley)	Gavage GD 14-18	Fetal T prod	G	A	A	G	G	A	G	G	G	High
Wang et al. 2017	Mouse (ICR)	Diet GD 0-21; GD 0-PND 21	Postnatal and Adult T conc	G	G	G	A	A	A	G	A	P	Medium
Oishi and Hiraga 1980a	Mouse (JCL:ICR)	Diet PND 35-42	Postnatal T conc	A	NR	NR	A	G	A	P	A	A	Medium
Oishi and Hiraga 1980b	Mouse (JCL:ICR)	Diet PND 35-42	Postnatal T conc	A	NR	NR	A	P	A	P	A	A	Medium
Hiraga 1980c	(JCL:Wistar)	PND 35-42	T conc										
Oishi and Hiraga 1980d	Rat (JCL:Wistar)	Diet PND 35-42	Postnatal T conc	A	NR	NR	A	P	A	A	A	A	Medium

Abbreviations: Gestation day (GD); Postnatal day (PND); Testosterone (T) production (prod) or concentration (conc)

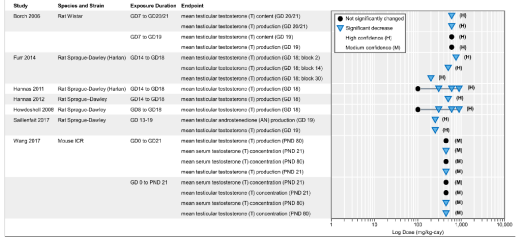


Figure 3. Summary of exposure-response for testosterone from gestational exposure studies.

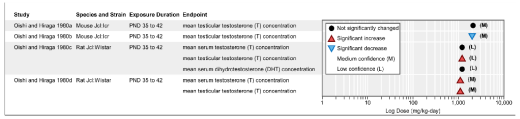


Figure 4. Summary of exposure-response for testosterone from postnatal exposure studies.

The synthesis of results for testosterone is summarized in an evidence profile table (Table 2). Gestational exposure studies provided *robust* evidence for effects on testosterone, whereas evidence from postnatal exposure studies was found to be *indeterminate*. Evidence judgments for other male reproductive endpoints identified in this systematic review are summarized in Table 3.

Table 2. Evidence profile table for animal studies of testosterone and DIBP

	Studies and interpretation	Factors that increase strength	Factors that decrease strength	Summary of findings
Gestational exposure	**High confidence** Borch et al. 2006 Furr et al. 2014 Hannas et al. 2011 Hannas et al. 2012 Howdeshell et al. 2008 Saillenfait et al. 2017 **Medium confidence** Wang et al. 2017	• Consistency • Exposure-response gradient • Effect size • Biological plausibility (support from mechanistic evidence) • Minimal concern for bias		⊕⊕⊕ ROBUST A dose-related decrease in testicular androgen levels or production (up to ~96% compared to control) was observed in all studies in rats and mice that evaluated this endpoint. Several of these studies also demonstrate decreased testicular expression of genes and proteins in the steroidogenesis pathway, which provides support for biological plausibility.
Postnatal exposure	**Medium confidence** Oishi and Hiraga 1980a Oishi and Hiraga 1980b Oishi and Hiraga 1980d **Low confidence** Oishi and Hiraga 1980c	• Biological plausibility	• High risk of bias • Unexplained inconsistency	◯◯◯ INDETERMINATE A dose-related increase in androgen levels was observed in two rat studies (Oishi and Hiraga 1980c-d), whereas androgen levels were decreased or not changed in mice (Oishi and Hiraga 1980a-b).

reproductive toxicity following DIBP exposure

Outcome	Includes these endpoints	Evidence following gestational exposure	Evidence following postnatal exposure
Testosterone	Androgen levels	Robust	Indeterminate
Male morphological development	AGD, nipple retention, preputial separation, hypospadias, cleft prepuce, exposed os penis, cryptorchidism	Robust	N/A
Sperm evaluation and histopathological effects in testis or epididymis	Sperm concentration and motility, oligospermia, azoospermia, granulomatous inflammation, tubular degeneration, tubular necrosis, interstitial hyperplasia	Robust	Moderate
Reproductive organ weight	Testis, epididymis, seminal vesicle weights	Moderate	Moderate
Male reproductive overall		Robust	

Discussion

Overall, the results from animal studies of male reproductive effects provide robust evidence of a hazard from DIBP exposure. Conclusions for testosterone are consistent with those of NAS (2017). The NAS review was limited to gestational exposure studies and excluded studies that exposed animals to a single high dose (≥500 mg/kg-day); therefore, NAS only considered two fetal testosterone studies, and had inadequate evidence to evaluate the effects of DIBP on AGD or hypospadias. The IRIS systematic review included all dose levels and life stages of exposure, and was able to evaluate a wider range of androgen-dependent and -independent male reproductive outcomes. *Disclaimer: The views expressed in this poster are those of the author(s) and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.*

Male reproductive toxicity in epidemiology studies of phthalates: a case study application of systematic review approaches

Elizabeth Radke, Glinda Cooper

U.S. EPA, Office of Research and Development (ORD), National Center for Environmental Assessment (NCEA), IRIS Division

Introduction

Phthalates have anti-androgenic activity in rodents resulting in reduced circulating testosterone and male reproductive tract abnormalities. Several epidemiologic studies have examined this association in humans. The National Academies of Sciences (NAS) recently published a systematic review of endocrine-related low-dose toxicity that included examination of phthalates and male reproductive tract development, and the Integrated Risk Information System (IRIS) performed a systematic review of all male reproductive effects of phthalate exposure, following recommendations in the 2014 NAS review of the IRIS program. Here, we use the associations between anogenital distance (AGD) in humans and two phthalates, di(2-ethylhexl phthalate (DEHP) and diisobutyl phthalate (DIBP), as a case study of the IRIS systematic review process. We also compare our conclusions to those of the NAS and summarize our overall findings on epidemiology studies of male reproductive effects of phthalates.

Methods

Epidemiology studies were identified by conducting a single broad literature search on the six phthalates of interest. The following databases were searched: PubMed, Web of Science, and Toxline. The last update was in January 2017. Title/abstract and full text screening was performed by two reviewers. Studies were evaluated by at least two reviewers using the approach in Figure 1.

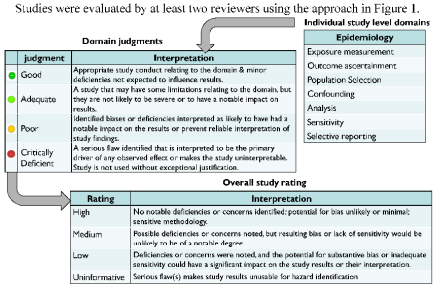


Figure 1. Study evaluation process

After study evaluation, the evidence for each outcome was synthesized for each phthalate, considering aspects of an association that may suggest causation. Based on this, the evidence was assigned within stream confidence judgments of *robust, moderate, slight, indeterminate,* or *compelling evidence of no effect.* The judgments for individual outcomes were summarized into an overall conclusion for male reproductive effects using a structured framework (see Poster by Yost et al.).

Figure 2. Abbreviated literature flow diagram

Results

Table 1. Epidemiology studies of AGD and phthalate exposure

	Reference	Study description: Population	Exposure	Outcome	Study evaluation: Exposure	Outcome	Selection	Confounding	Analysis	Overall confidence
Included	Bornehag et al., 2015	Birth cohort (N=196 boys) in Sweden	Single urine sample (1st trimester)	AGD at 19-21 mo	A/P	G	G	A	G	Medium
	Bustamante-Montes et al., 2013	Birth cohort (N=73 boys) in Mexico	Single urine sample (3rd trimester)	AGD at 1-2 d	P	G	A	A	G	Low
	Jensen et al., 2016	Birth cohort (N=273 boys) in Denmark	Single urine sample (26-30 wk gestation)	AGD at 3 mo	A/P	G	G	A	G	Medium
	Suzuki et al., 2012	Birth cohort (N=73 boys) in Japan	Single urine sample (3rd trimester)	AGD at 1-3 d	P	A	P	P	A	Low
	Swan, 2008	Birth cohort (N=106 boys) in U.S.	Single urine sample (mean 28 wk gestation)	AGD at 1-36 mo	A/P	P	G	A	P	Low
	Swan et al., 2015	Birth cohort (N=365 boys) in U.S.	Single urine sample(1st trimester)	AGD at 1-2 d	A/P	G	A	A	A	Medium

G=good, A=adequate, P=poor, A/P=adequate for short chain phthalates, poor for long chain.
Studies with biomarker measures based on samples other than urine (e.g., blood) were considered to be critically deficient for all short-chain phthalates and for primary metabolites (e.g., MEHP, MINP) of long-chain phthalates.

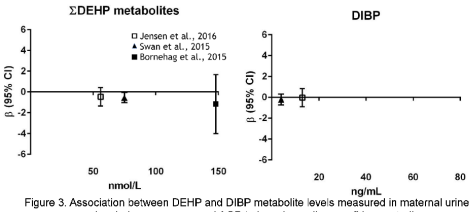


Figure 3. Association between DEHP and DIBP metabolite levels measured in maternal urine samples during pregnancy and AGD in boys in medium confidence studies

Regression coefficients on the y-axis are plotted against exposure level on the x-axis (population median for each study).

Table 2. Evidence profile table for epidemiology studies of AGD and DEHP and DIBP

	Studies and interpretation	Factors that increase strength	Factors that decrease strength	Summary of findings and within stream evidence judgment
DEHP	**Medium confidence** Bornehag et al., 2015 Jensen et al., 2016 Swan et al., 2015 **Low confidence** Bustamante-Montes et al., 2013 Suzuki et al., 2012 Swan, 2008	Among medium confidence studies: • consistency • exposure-response gradient across studies • minimal concerns for bias	• low precision in study with largest effect size	⊕⊕◯ MODERATE Inverse associations between DEHP exposure and anogenital distance reported in 5/6 studies (Jensen et al., 2016, Swan et al., 2015, Bornehag et al., 2015, Swan, 2008, Suzuki et al., 2012), of which 2 were statistically significant (Swan et al., 2015, Swan, 2008). Among the 3 medium confidence studies, effect size increased with increasing exposure levels.
DIBP	**Medium confidence** Jensen et al., 2016 Swan et al., 2015 **Low confidence** Swan, 2008	• low study sensitivity may explain lack of association	• inconsistency • few studies	⊕◯◯ SLIGHT Inverse associations between DIBP exposure and anogenital distance reported in 2/3 studies (Swan, 2008, Swan et al., 2015), though neither were statistically significant. Exposure levels and range were low in all studies.

Of the seven identified studies on phthalates and AGD (Figure 2), one was excluded due to inadequate exposure measurement. Summary of the evaluations for the six included studies is in Table 1. Results of medium confidence studies were given priority (Figure 3), but all studies were included in the synthesis, which is summarized in the evidence profile table (Table 2). For DEHP, an exposure response gradient was observed across studies, with the study with the highest exposure levels reporting the strongest association. This was not observed for DIBP, but exposure levels were low in all studies. The same methods were used for other phthalate/outcome combinations and the within stream evidence judgments are shown in Figure 4. Table 3 presents a comparison of the within stream judgments from the IRIS and NAS reviews of anogenital distance, testosterone in infants, and hypospadias. Both found that the evidence for the latter two outcomes was not adequate to form a conclusion. For anogenital distance, evidence for DEHP and DBP was considered *moderate* in both reviews. Evidence for DINP, DIBP, and BBP was considered *slight* by IRIS and *inadequate* by NAS. These conclusions were not considered inconsistent, but rather reflect differences in the process for evidence synthesis. Only DEP differed between reviews, classified as *slight* by IRIS and *moderate* by NAS based on the results of a meta-analysis.

Outcome	DEHP	DINP	DBP	DIBP	BBP	DEP
Anogenital distance	M	S	M	S	S	S
Hypospadias/cryptorchidism	I	S	S	S	S	I
Pubertal development	S	S	S	S	S	S
Semen parameters	M	M	R	S	M	S
Time to pregnancy	S	I	M	S	M	I
Testosterone	M	M	S	M	I	I
Male repro overall	R	M	R	M	M	S

Robust (R) | Moderate (M) | Slight (S) | Indeterminate (I)

Figure 4. Within stream evidence judgments for human evidence of male reproductive effects associated with phthalates

Table 3. Within stream evidence judgments of systematic reviews of male reproductive developmental toxicity in epidemiology studies by IRIS and NAS

	Anogenital distance		Testosterone in infants		Hypospadias	
Phthalate	IRIS	NAS	IRIS	NAS	IRIS	NAS
DEHP	Moderate	Moderate	Indeterminate	Inadequate	Indeterminate	Inadequate
DINP	Slight	Inadequate	Indeterminate	Inadequate	Indeterminate	Inadequate
DBP	Moderate	Moderate	Indeterminate	Inadequate	Indeterminate	Inadequate
DIBP	Slight	Inadequate	Indeterminate	Inadequate	Indeterminate	Inadequate
BBP	Slight	Inadequate	Indeterminate	Inadequate	Indeterminate	Inadequate
DEP	Slight	Moderate	Indeterminate	Inadequate	Indeterminate	Inadequate

Classifying levels: IRIS: Robust, Moderate, Slight, or Indeterminate; NAS: High, Moderate, Low, or Inadequate

Discussion

Overall, the results from epidemiology studies of male reproductive effects provide evidence of a hazard from phthalate exposure. Looking specifically at anogenital distance, there is *moderate* evidence of an association with DEHP and DBP exposure, and *slight* evidence for other phthalates. These findings are generally consistent with the NAS report on low-dose toxicity from endocrine active chemicals (2017). In the case of DIBP, the weaker evidence may be largely explained by the smaller number of studies and low exposure levels that decreased study sensitivity.

U.S. Environmental Protection Agency
Office of Research and Development

Quantitative Evaluation of Uncertainty: APROBA and Beyond

Todd Blessinger and David Bussard

US EPA, Office or Research and Development, National Center for Environmental Assessment - Washington

Todd Blessinger | Blessinger.todd@epa.gov | 202-564-7641

Purpose and Scope

Quantitative assessment of uncertainty was recommended by the NRC:

- *Science and Decisions* report (NRC, 2009) – recommended incorporating probabilistic methods for assessing uncertainty.
- *Review of the IRIS Program* report (NRC, 2014) – recommended systematic use of uncertainty analysis and expanded use of Bayesian methods.

NCEA will pilot this approach to better understand issues in implementing it and to engage in dialogue with stakeholders as to advantages and challenges in utilizing this approach.

Probabilistic Calculation of Risk-Specific Doses

Goal: Probabilistically incorporate adjustments and uncertainty when extrapolating dose-response results from animal data to the human population.

Current Practice: Reference values (RfVs) are generally calculated by dividing a point of departure (POD; usually a BMDL or NOAEL) by a series of uncertainty factors (UFs):

$$\text{RfV} = \frac{\text{POD}}{\text{UF}_1 \times \cdots \times \text{UF}_k}$$

- Default values of UFs are (1, 3, or 10).
- Decision on which value to use is made qualitatively based on information available for the particular assessment (e.g., size of database, study characteristics)
- Reference Value definition does not explicitly target incidence, effect size, or confidence.

Proposed New Practice: Calculate risk-specific dose intervals using probabilistically-defined versions of POD and UFs, using the concept of target human dose.

Target Human Dose and APROBA

Target human dose, HD_M^I:

- HD_M^I = the Human Dose at which a fraction (or incidence) I of the population shows an effect of magnitude (or severity) M or greater for the critical effect considered.
  - A "risk-specific dose."

Examples:

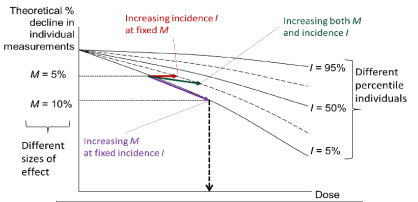
- HD_{10}^{01} – human dose at which 1% of the population shows an increase in liver weight of 10% or greater above background.
- HD_{05}^{01} – human dose at which there is an individual extra risk of lung tumors of 5% (or more) in 1% of the population.

HD_M^I is calculated using the formula similar to RfV:

$$HD_M^I = \frac{\text{POD}}{AF_1 \times \cdots \times AF_k} \quad (1)$$

- Each AF, or "assessment factor," is treated as a continuous random variable; the parameters of the distributions of these random variables can be determined from empirical data. The resulting HD_M^I is a random variable with its own probability distribution.

Target Human Dose (cont'd)



Key Concept: HD_M^I (e.g., HD_{10}^{05})

Approximate Probability Analysis (APROBA) is an Excel-based tool to calculate a probabilistic RfV from animal data.

- Computes HD_M^I under the assumption that the POD and AFs are independent lognormally distributed.
- An analogue to a reference value can be derived for a pre-selected percentile (e.g., 5th percentile) of the HD_M^I distribution. The interval reflects uncertainty as well as a choice of a desired confidence (e.g., 95%) in the HD_M^I estimate.
- Was applied by the Dutch National Institute for Public Health and the Environment (RIVM) in recent risk assessment on melamine.

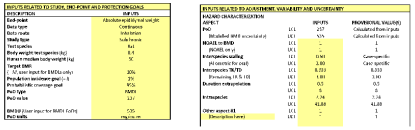
Example

Dose-response data of absolute epididymis weight in adult rats after exposure to chemical X by inhalation:

Exposure (ppm)	No. of animals	Mean (mg)	SD (mg)
0	25	0.3327	0.03631
100	25	0.3311	0.04453
250	25	0.3053	0.04188
500	25	0.2912	0.05206
750	25	0.2405	0.04804

Exponential model 3 fit to data at BMR of 10% relative deviation from control mean yields: BMDL = 237 ppm; BMDU = 535 ppm

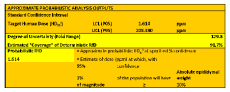
Input in APROBA worksheet:



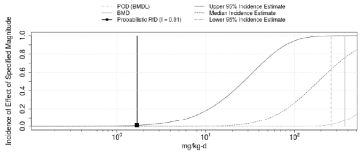
- Input on left entered by user
- Values on right are lower and upper confidence limits representing the estimated 5th and 95th percentiles of the lognormal distribution for the AFs.
  - LCL and UCL calculated using empirical data
- HDMI has lognormal distribution based on formula in Equation (1).

Example (cont'd)

APROBA output:



RfV = 1.6 ppm, which is the LCL (P05 = 5th percentile) of the HDMI distribution.



Plot: CDFs of Lower, Median, and Upper Incidence Estimates

- Several types of "central" estimates can be derived, such as the median or the expected value, if assuming a log-normal distribution.
- The approach could also be modified to provide a distribution on the population risk at a given dose.
  - Distribution can be used to estimate benefits of reduced exposures or for communication about risks of exposure.

Next Steps

- Conduct a case study using APROBA to evaluate the advantages of incorporating quantitative uncertainty in assessments with this approach.
  - Evaluate the information and choices needed to produce the estimates.
  - Work with risk managers to evaluate if this approach is useful, and how it might need modification to be more useful.
- Apply uncertainty analysis to risk assessment done to support benefit-cost analysis.
- Non-APROBA-based uncertainty analysis.

References

- IPCS (International Programme on Chemical Safety). (2014). Guidance document on evaluating and expressing uncertainty in hazard characterization. World Health Organization.
- Chiu, WA; Slob, W. (2015). A unified probabilistic framework for dose–response assessment of human health effects. Environ Health Perspect (123) pp 1241-1254.
- Risk assessment and derivation of a provisional guideline value for melamine in drinking water. Advice to: Ministry of Infrastructure and Environment (Inspectorate of Environment and Transport) RIVM.

U.S. Environmental Protection Agency
Office of Research and Development

Printed on 100% recycled/recyclable paper with a minimum 50% post-consumer fiber using vegetable-based ink.

EPA United States Environmental Protection Agency

Mode of action and human relevance evaluation of Dibutyl Phthalate (DBP)-induced male reproductive system toxicity.

Xabier Arzuaga[1]; Teneille Walker[1]; Andrew Hotchkiss[1]
[1]US EPA, Office of Research and Development, National Center for Environmental Assessment

Introduction

Dibutyl phthalate (DBP) is used as a plasticizer in a variety of commercial and consumer products (US EPA, 2014; Kavlock et al. 2002). The largest source of DBP exposure in humans is food, with inhalation and dermal exposures considered minimal (Kavlock et al. 2002). Epidemiological studies provide evidence of human exposure and altered androgen levels during lifestages at which androgen production is critical for the normal development and function of the male reproductive system (WHO/UNEP, 2013), and experimental studies using rat models have reported that exposure to DBP is associated with adverse responses in the male reproductive system. Effects include decreased androgen production, agenesis of the male reproductive system and increased incidence of internal and external malformations after developmental exposures (e.g. degeneration of seminiferous tubules, hypospadias), and decreased fertility and sperm counts (CPSC, 2010; Makris et al 2013; US EPA, 2009). Evidence from post-natal exposure studies also suggests that young animals are more sensitive to phthalate-induced testicular injury than adults (Boekelheide et al 2004). However, recent studies using ex-vivo human tissue culture preparations, or rodent and human testicular tissue xenografts report that human fetal testes are resistant to phthalate induced disruption of testosterone production (Johnson et al., 2012; Albert and Jégou, 2014). Such findings raise questions about the human relevance of the androgen-related endpoints measured in experimental rodents exposed to phthalates.

A mode of action framework was used to evaluate the available evidence from experimental and in-vitro studies according to lifestage of exposure. Studies considered for this analysis include:

- Exposures during the masculinization programming window (MPW; gestational period during which development of the male reproductive system occurs).
- Exposures during early post-natal stages.

Methods

The experimental and mechanistic studies considered in this analysis were obtained from the literature search performed by the US EPA Integrated Risk Information System (IRIS). Studies for DBP or MBP were identified from online databases (PubMed, Web of Science, Toxline, and TSCATS2) using search terms designed to capture pertinent studies. The last update was performed in July 2017. Title/abstract screening followed by a full text review was performed to identify relevant studies on male reproductive effects and related mechanisms/pathways (See Figure 1 below). The types of in-vivo and in-vitro studies considered most informative to our evaluation were:

- Gestational DBP exposure studies that use mammalian in-vivo and in-vitro models, and human xenograft and ex vivo models treated during the masculinization programming window.
- Additional ex-vivo studies that expose human fetal testis tissue cultures to DEHP or its metabolite MEHP.
- Studies aimed at characterizing the receptor for DBP at a molecular level.
- Post-natal DBP exposure studies that use mammalian model species, including in-vivo, xenograft, and cell culture models.

The available mechanistic and toxicological evidence was analyzed in concordance with the framework and levels of biological organization used for mode of action Action analysis for non-cancer effects and development of Adverse Outcome Pathways (Bobbis et al 2008; Edwards et al 2016). As recommended by US EPA's Framework for Assessing Health Risk of Environmental Exposures to Children and the World Health Organization International Programme on Chemical Safety, the available mechanistic and toxicological studies and endpoints that inform the mode of action for DBP-induced male reproductive effects were evaluated according to the lifestage of exposure

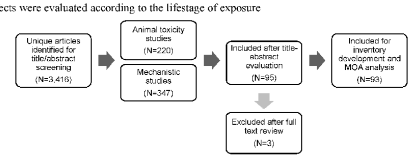


Figure 1. Abbreviated literature flow diagram

Disclaimer: The views expressed in this poster are those of the author(s) and do not necessarily represent the views or the policies of the U.S. Environmental Protection Agency.

U.S. Environmental Protection Agency
Office of Research and Development

Figure 2: Pathway for DBP-induced male reproductive effects after gestational exposure during MPW

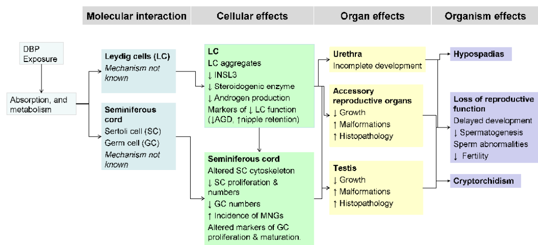
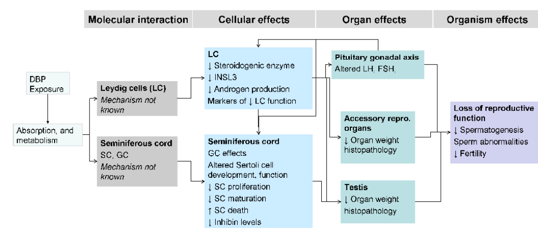


Figure 3: Pathway for DBP-induced male reproductive effects in post-natal lifestages



Results and discussion

Gestational exposure studies:

- Fetal rats appear more sensitive to DBP-induced anti-androgenic effects than are mice and may be more sensitive than other rodent species, non-human primates, and human fetal testis xenografts and ex-vivo tissue cultures.
- DBP-induced androgen-independent effects in the seminiferous cord (SC & GC) are conserved among most mammalian models (rats, rabbits and mice) and human xenografts.

Post-natal lifestage studies using peri-pubertal or sexually mature animals:

- DBP-induced Leydig cell effects are conserved in different mammalian species: (rats, rabbits, mice, gerbils, and guinea pigs, non-human primates [in-vivo and xenografts]).
- DBP-induced effects in the seminiferous cord (SC & GC) are also conserved among most mammalian models (rats, mice, and non-human primate [xenograft]).

Table 1: Preliminary cross-species coherence analysis for gestational effects

Key event	Animal in-vivo evidence	Animal (ex-vivo, xenograft)	Humans evidence (ex-vivo, xenograft)
Leydig cells (LCs)	No evidence		Not identified in studies
Sertoli cells (SCs), germ cells (GCs)	No evidence		Not identified in studies
LCs	■ Rat [*ms*] & rabbits [1] ■ Mice [1] ⊐ Marmosets [1] & mice [3]	■ Rat xenograft [2] □ Rat ex-vivo* [2] ⊔ Mice xenograft [1] ⊓ Mice ex-vivo [1]	⊐ Human xenografts [3] ⊐ Human ex-vivo [2]
SCs, GCs	■ SC and GC effects in rats [*ms*] ■ SC and/or GC effects in rabbits [1] mice [3] ⁻ Marmoset [1]	■ Mice ex-vivo [2] ■ Mice xenograft [1]	■ Human xenografts [3] ■ Human ex-vivo [1]
Urethra	■ Rats [*ms*]		Not evaluated
Accessory reproductive organs	■ Rats [*ms*] & rabbits [1] ⊐ Marmoset [1]	■ Rat xenograft [1]	¬ Human xenograft [2]
Testis	■ Rats [*ms*], rabbits [1] & mice [2] ⊐ Marmoset [1] & mice [3]		No evaluated
Organism effects: reproductive functions	■ Rats [*ms*] & rabbits [1] Marmoset [1]		Not evaluated

■ – Evidence of response to exposure
⊐ – Evidence of no response (or reduced sensitivity) to exposure
[ ] – number of studies identified in the literature
ms – Many studies
* Both rat and mouse ex-vivo studies report no effect on basal T production, but gonadotropin-stimulated T was inhibited by exposure.

Table 2: Preliminary cross-species coherence analysis for effects in early post-natal lifestages

Key Event	Animal evidence (in-vivo)	Animal evidence (cell culture, xenograft)	Human evidence (ex-vivo, xenograft)
Leydig cells (LCs)	No evidence		No studies available
Sertoli cells (SCs), Germ cells (GCs)	No evidence		No studies available
LCs	■ Rats [17], mice [3], rabbits [1], marmoset [1], ⊐ Rats [1], mice [1]	■ Cell culture models (rat [3], mouse [7] & dog [1]). ■ Rhesus monkey xenografts [1]	No studies available
SCs, GCs	■ Rats [30], mice [5], ⊐ Marmoset [1]	■ Cell culture (rats [9]), mice [3], rhesus monkey xenografts [1]	No studies available
Pituitary gonadal axis	■ Rats [6] ¬ Rabbits [1], mice [3], rats [1]		No studies available
Accessory reproductive organs	■ Rats [10], rabbits [1], gerbils [1], mice [2] ⊐ Mice [3], rats [8]	■ Rhesus monkey xenografts [1]	No studies available
Testis	■ rats [34], rabbits [1], mice [5], & guinea pigs [1] ¬ Hamsters [1], rats [5] mice [3], marmosets [2]	■ Rhesus monkey xenografts [1]	No studies available
Reproductive functions	■ Rats [12], rabbits [1], mice [5], & guinea pigs [1] ⊐ mouse [2], rats [1]		No studies available

■ – Evidence of response to exposure
⊐ – Evidence of no response (or reduced sensitivity) to exposure
[ ] – number of studies identified in the literature

Selected references

Albert O, and Jégou B, *Hum Reprod Update*. 2014. 20(2): 231-49.
Edwards et al., *Journal of Pharmacology and Experimental Therapeutics*. 2016. 356(1): 170-181.
Howdeshell KL, et al., *Environ Res*. 2008. 108(2): 168-76.
Makris SL, et al., *Birth Defects Res B Dev Reprod Toxicol*. 2008. 83(6): 530-46.
Boobis AR, et al., *Crit Rev Toxicol*. 2008. 38(2): 87-96.
Johnson KJ. et al., *Toxicol Sci*. 2012. 129(2): 235-48.
US EPA. (2009) An approach to using toxicogenomic data in US EPA human health risk assessments: a dibutyl phthalate case study.
US EPA. (2006) A framework for assessing health risk of environmental exposures to children.

Printed on 100% recycled/recyclable paper with a minimum 50% post-consumer fiber using vegetable-based ink.

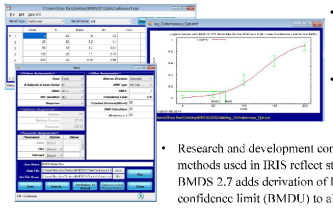
EPA Dose-Response & Related Software - New & Future Developments

United States Environmental Protection Agency

J. Allen Davis[1], Jeff Gift[2], David Farrar[1], Jay Zhao[1], and Matt Wheeler[3]

[1] US EPA, Office of Research and Development, National Center for Environmental Assessment - Cincinnati
[2] US EPA, Office of Research and Development, National Center for Environmental Assessment – Washington
[3] National Institute for Occupational Safety and Health, Risk Evaluation Branch – Cincinnati, OH

Benchmark Dose Software (BMDS 2.7 released 8/17)

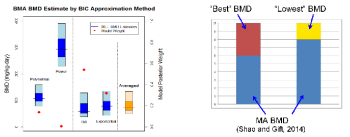


- Benchmark dose (BMD) method proposed by Crump (1984)
- Accepted as default dose-response modeling approach by US EPA (2012)
- Research and development continues to ensure methods used in IRIS reflect state-of-the-science, e.g., BMDS 2.7 adds derivation of BMD upper bound confidence limit (BMDU) to all models (USEPA 2017)

BMDS 3.0 - to be released in FY18

Bayesian Model Averaging

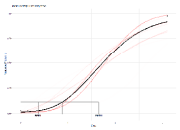
- EPA NCEA and NIOSH are developing Bayesian modeling averaging methods to address and/or account for model uncertainty
- Current methods for single model selection (i.e., AIC-based selection) have been shown to be inadequate (i.e., methods do not achieve nominal coverage rates)
- Current method uses maximum a posteriori estimation and Laplace approximations to generate model weights



MA may reduce uncertainty in BMD estimates

MA may enhance accuracy of BMD estimates

- Method allows for assignment of model parameters and model weights, allowing for incorporation of biological or other prior information
- For example, information of a particular endpoint's mode of action may support weighting non-linear models more heavily than linear ones



$$\Pr(BMD \mid D) = \sum_{i=1}^{9} \pi_i \Pr(BMD \mid M_i, D)$$

Posterior Distribution of the BMD

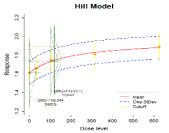
$$\alpha = \int_{-\infty}^{BMD_L} \Pr(BMD \mid D) dBMD$$

Calculation of the BMDL

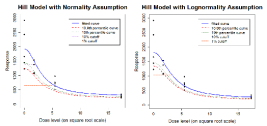
U.S. Environmental Protection Agency
Office of Research and Development

BMDS 3.0 - to be released in FY18 (continued)

- **Hybrid Approach** – instead of using change in central tendency, the hybrid approach estimates a BMD using the percentage change of a population in the tail of the distribution
- Use of the hybrid approach for continuous data harmonizes benchmark responses between continuous and dichotomous data



Application of the hybrid approach to estimate BMDs for a continuous endpoint



Comparison of dose-response curves under the normal or log-normal distributional assumptions

- **Log-normality vs. Normality** – Shao and Gift (2013) determined that the distribution assumption has limited impact on the BMD estimates when the within dose-group variance is small
- BMDs defined using the hybrid approach are more sensitive to the distribution assumption

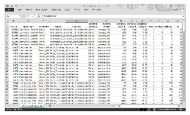
Categorical Regression (CatReg 3.1 released 6/17)

Categorical Regression

- Estimates the probability that a response occurs of a severity level, s, or greater given a concentration, C, and duration of exposure, T, as:

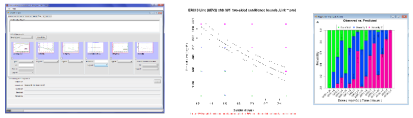
$$P(Y \geq s|C,T) = H[\alpha_s + \beta_{1s} * C + \beta_{2s} * T]$$

- CatReg allows for meta-analysis of data from multiple studies, endpoints, and test species (USEPA 2017; Milton et al., 2017)
- CatReg accounts for within study correlations (clustering) and allows for the stratification of model parameters to account for response differences across strata of data.



$$\Pr(Y \geq s|C,T,i) = H[\alpha_s + \gamma_i + \beta_1 j \times f_1(C) + \beta_2 k \times f_2(T)],$$
$$s = 1,2,\ldots,S,\quad i = 1,2,\ldots I,\quad j = 1,2,\ldots,J,\quad k = 1,2,\ldots,K$$

- CatReg incorporates hypothesis testing to allow users to determine the most appropriate form of the model (i.e, which variables should be stratified)
- Multiple plotting capabilities are implemented in CatReg

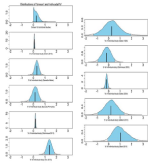


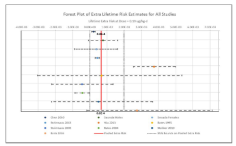
- U-shaped dose-response analysis could be added to future CatReg versions to facilitate assessment of toxicity from excess and deficiency (Milton et al., 2017)

Some Additional Related Developments and Plans

Probabilistic Meta-Analysis Methods for Meta-Analysis of Epidemiological Data

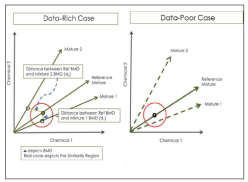
- Probabilistic meta-analysis dose-response methods have been proposed (NRC, 2008, 2013) to better assist risk management decision making
- Meta-analysis tools that allow for the combination of a multiple types of epidemiological studies using Bayesian statistics and hierarchical modeling have been developed to support future Agency health assessments





Mixture Similarity Tool (MiST)

- EPA Excel tool (MiST) based on Marshall et al. (2013)
- Data-Rich Case: Mixtures are similar when distance between reference and candidate mixture BMDs is less than radius of red circle
- Data-Poor Case: Simplifying assumptions to estimate distance via comparison of mixing proportions and weights for components of reference & candidate mixtures.



Addressing NRC Recommendations

New and future developments in dose-response modeling specifically address multiple recommendations provided by NRC (2014)

- "EPA should use formal methods for combining multiple studies and the derivation of IRIS toxicity values"
  - Both CatReg and meta-analysis tools for epidemiological data have been developed to increase IRIS' meta-analytical capabilities
- "Advanced analytic methods, such as Bayesian methods, for integrating data from dose-response assessments and deriving toxicity estimates are underused in the IRS program
  - Bayesian methods have recently been developed for use in IRIS assessments, including Bayesian model averaging and hierarchical Bayesian meta-regression approaches
- "Uncertainty analysis should be conducted systematically and coherently in IRIS assessments
  - Uncertainty analysis is supported by reporting entire confidence interval around BMD (BMDL – BMDU), which is done in the new model averaging method and CatReg

References

Marshall et al. (2013) An empirical approach to sufficient similarity: combining exposure data and mixtures toxicology data. *Risk Analysis*, 33(9), pp. 1582-1595

Milton et al. (2017) Modeling U-shaped dose-response curves for manganese using categorical regression. *Neurotoxicology*, 58, 217-225

Shao, K. and Gift, J.S., 2014. Model uncertainty and Bayesian model averaged benchmark dose estimation for continuous data. *Risk Analysis*, 34(1), pp.101-120.

US EPA, 2012, Benchmark Dose Technical Guidance Document, https://www.epa.gov/risk/benchmark-dose-technical-guidance

US EPA 2017a Benchmark Dose Software (BMDS) v. 2.7, https://www.epa.gov/bmds

US EPA 2017b Categorical Regression (CatReg) v. 3.1, https://www.epa.gov/bmds/catreg

Evidence profile table for DIBP and male reproductive toxicity

Erin E. Yost[1], Xabier Arzuaga[2], Elizabeth Radke[2]

[1] US EPA National Center for Environmental Assessment, Research Triangle Park, NC; [2] US EPA National Center for Environmental Assessment, Washington, DC

The evidence profile table is a tool that complements the evidence integration narrative for human and animal data. Explanations for factors that increase or decrease confidence are provided in summaries.

Outcome*		Studies and interpretation	Factors that increase strength	Factors that decrease strength	Summary of findings	Within stream evidence judgment	Inference across evidence streams	Overall conclusion
HUMAN STUDIES							**Relevance of animal data to humans** •Role of androgen-dependent and –independent pathways in male reproductive system development, maturation, and function is conserved across mammalian species. **Cross-stream coherence** •Testosterone is reduced with phthalate exposure in both humans and animals during different lifestages. **Susceptibility** •Developmental stages are particularly susceptible to perturbation by phthalates **Other relevant information** •Evidence from DBP, a structurally similar phthalate, provides robust evidence of male reproductive toxicity in humans, likely due to higher exposure levels and a larger number of studies	⊕⊕⊖ Overall conclusion that DIBP causes male reproductive toxicity, based on: 1) Robust evidence from oral exposure studies in rats and mice, with significant outcomes in gestational exposure studies at doses as low as 300 mg/kg-day; 2) Moderate evidence in human epidemiological studies of decreased testosterone in adult men with median metabolite concentrations in urine ranging from 7-48 ng/mL. Evidence for other outcomes was from populations with low urine metabolite concentrations, which reduced study sensitivity; and 3) Supporting mechanistic evidence demonstrating decreased testicular steroidogenesis and INSL-3. Evidence from animals is presumed relevant to humans. Lower level of evidence in humans can be explained by low sensitivity and few available studies.
Testosterone (adult)		(All cross sectional studies) **Medium confidence** Meeker and Ferguson (2014) Pan et al., 2015 **Low confidence** Chang et al. (2015) Den Hond et al. (2015)	• Consistency • Minimal risk of bias in medium confidence studies	• Few studies available	⊕⊕◯ MODERATE Inverse associations between DIBP exposure and testosterone levels in 3/4 studies (Meeker and Ferguson et al., 2014, Pan et al., 2015, Chang et al., 2015), 2 of which were statistically significant. No studies examined exposure-response gradient.	⊕⊕◯ MODERATE Based on data for testosterone in adults, supported by slight evidence in other outcomes with low sensitivity and few available studies explaining lack of clear associations.		
Anogenital distance (AGD), semen parameters, pubertal development, time to pregnancy, hypospadias/cryptorchidism					⊕◯◯ SLIGHT			
ANIMAL STUDIES								
Gestational exposure	Testosterone	**High confidence** Borch et al. 2006 Furr et al. 2014 Hannas et al. 2011 Hannas et al. 2012 Howdeshell et al. 2008 Saillenfait et al. 2017 **Medium confidence** Wang et al. 2017	• Consistency • Exposure-response gradient • Effect size • Biological plausibility (support from mechanistic evidence) • Minimal risk of bias		⊕⊕⊕ ROBUST A dose-related decrease in testicular androgen levels or production (up to ~96% compared to control) was observed in all studies in rats and mice that evaluated this endpoint. Several of these studies also demonstrate decreased testicular expression of genes and proteins in the steroidogenesis pathway, which provides support for biological plausibility.	⊕⊕⊕ ROBUST Supported by consistency and coherence across outcomes, with mechanistic evidence (e.g. decreased testicular expression of steroidogenic enzymes and INSL-3) providing support for biological plausibility. The greatest weight of evidence came from gestational exposure studies, whereas postnatal exposure studies were limited by risk of bias concerns.		
	Male morphological development	**High confidence** Borch et al. 2006 Saillenfait et al. 2006 Saillenfait et al. 2008 Saillenfait et al. 2017 **Medium confidence** Wang et al. 2017	• Consistency within rat studies • Exposure-response gradient • Effect size • Biological plausibility • Minimal risk of bias		⊕⊕⊕ ROBUST All rat studies observed a dose-related increase in effects consistent with decreased testosterone and INSL-3, including increased time to puberty, decreased AGD, nipple retention, cryptorchidism, hypospadias, exposed os penis, and cleft prepuce. No effects on AGD were observed in mice (Wang et al. 2017).			
	Sperm evaluation and histopathological effects in testis or epididymis	**High confidence** Saillenfait et al. 2008 **Medium confidence** Borch et al. 2006 Wang et al. 2017	• Consistency • Exposure-response gradient • Effect size • Biological plausibility		⊕⊕⊕ ROBUST Adverse effects on the testis and/or sperm were observed in rats and mice, including a dose-related increased incidence of pathological lesions of the testis (Borch et al. 2006, Saillenfait et al., 2008), epididymal oligo- or azoospermia (Saillenfait et al. 2008), and decreased sperm concentration and motility (Wang et al. 2017).			
	Reproductive organ weight	**High confidence** Saillenfait et al. 2008 **Medium confidence** Wang et al. 2017	• Exposure-response gradient • Biological plausibility • Minimal risk of bias	• Few studies	⊕⊕◯ MODERATE Decreased reproductive organ weights were observed in rats (Saillenfait et al. 2008), whereas a consistent trend in testis weight was not observed in mice (Wang et al. 2017).			
Postnatal exposure	Testosterone				◯◯◯ INDETERMINATE			
	Sperm evaluation and histopathological effects in testis or epididymis	**Low confidence** Oishi and Hiraga 1980c Foster et al. 1981	• Consistency • Biological plausibility • Coherence with gestational exposure studies	• High risk of bias • Few studies	⊕⊕◯ MODERATE Rats were found to have increase testicular atrophy (Foster et al. 1981) and decreased spermatocytes and spermatogonia (Oishi and Hiraga 1980c).			
	Reproductive organ weight	**Medium confidence** Oishi and Hiraga 1980a Oishi and Hiraga 1980b Oishi and Hiraga 1980c Oishi and Hiraga 1980d **Low confidence** Foster et al. 1981 U. Rochester 1954 Zhu et al. 2010	• Consistency within rat studies • Biological plausibility • Coherence with gestational exposure studies	• High risk of bias in some studies • Unexplained inconsistency	⊕⊕◯ MODERATE In rats, a dose-related decrease in absolute testis weight was consistently observed (Oishi and Hiraga 1980c-d, Foster at al. 1981, University of Rochester 1954). In mice, Zhu et al. (2010) observed decreased testis weight in the highest dose group, whereas Oishi and Hiraga (1980a-b) observed increased testis weight.			

*Outcomes with slight or indeterminate evidence received a full systematic review, but were not significant contributors to the overall conclusion, so the details of the evidence are not provided here.

Comparison with findings from the National Academy of Sciences (NAS) systematic review of the low-dose toxicity of phthalates (2017)

Table 1: Summary of conclusions for DIBP from NAS 2017.

Endpoint	Initial hazard evaluation
Testosterone	Presumed human hazard • Based on high level of evidence from animal studies and inadequate evidence from human studies
AGD	Not classifiable • Based on inadequate evidence from human and animal studies
Hypospadias	Not classifiable • Based on inadequate evidence from human and animal studies

Both IRIS and NAS (2017) concluded that DIBP is likely to cause male reproductive toxicity in humans. However:

- NAS was only able to draw this conclusion for testosterone, based on the high level of evidence from rodent studies. Other endpoints (AGD and hypospadias) were determined to have inadequate evidence available.
- The IRIS systematic review was broader in scope (see Table 2) and was able to draw conclusions for a range of androgen-dependent and –independent male reproductive outcomes.

Table 2: Summary of major scoping differences between the IRIS and NAS systematic reviews of DIBP

	IRIS	NAS
Exposure	All life stages and dose levels	In utero exposure; Animal studies using a single dose ≥500 mg/kg-day are excluded.
Outcomes	Any male reproductive outcome	Testosterone, AGD, hypospadias

U.S. Environmental Protection Agency
Office of Research and Development

A New Bayesian Approach to Combining Different Species Data

Leonid Kopylev and Junyong Park

US EPA, Office of Research and Development, National Center for Environmental Assessment – Washington

University of Maryland at Baltimore County, Department of Mathematics and Statistics and ORISE Faculty Fellow at US EPA/ORD/NCEA-Washington

Purpose and Scope

NRC (2014) recommended that IRIS develop the capacity to do Bayesian modeling of chemical hazards. In particular, NRC stated that "...more sophisticated Bayesian approaches have been proposed for combining dose-response estimates for multiple species and multiple chemicals (DuMouchel and Harris 1983; Jones et al. 2009). Those approaches might also be useful to EPA if guidance for selection of appropriate models and priors is developed."

In this research and development effort, EPA evaluated DuMouchel and Harris (1983) approach, developed alternative approach and applied it to the data in Jones et al. 2009

Background

DuMouchel and Harris (1983, JASA) were the first authors that addressed the problem of combining information for multiple species with a non-simplistic approach.

- Proposed a Bayesian approach to interspecies extrapolation
- Special attention to combining dose-response information
- Realized that they need subject matter expertise

DuMouchel and Harris (1983) realized that

- the species do not need to be restricted to humans and animals
- any type of data (including cell potency) is appropriate.
- a lot of toxicological experience is needed to figure out what chemicals dose-response information is combined.

Their ANOVA structure, however, assumes constant relative potency across species, which may not be the case in many examples.

A suggested model: Gaussian graphical model

Example (Jones et al. 2009; Low birth weight)

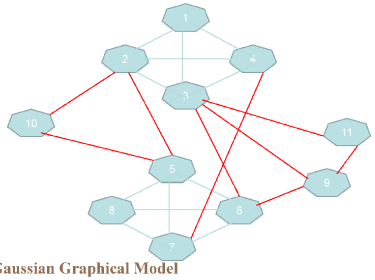
	Total THMs	Chloroform	BDCM	DBCM	Bromoform
Humans	1	2	3	4	12 (missing)
Rat S-D		5	6	7	8
Rat S-F			9		
Rabbit D-B		10			
Rabbit N-Z			11		

- Cell 1 - Cell 11 have the slope of regression model from log(dose) and log(response)
- Empty cells represent no-data

Assumptions

- We assume that species are related for the same chemical and different chemicals are related for the same species.
- We model dependence or relationship among different species and different chemicals through edges in Gaussian graphical model.
- We need to control dependence through prior probabilities for edges based on scientific knowledge rather than subjective choice.

Graphical representation of the Example



Gaussian Graphical Model

- Gaussian graphical model uses the inverse of covariance matrix called a precision matrix.
- Each component in a precision matrix represents the partial correlation between two nodes in the graphical model. Red edges shown for the same chemical data.
- No-edge between two nodes in graphical model is equivalent to partial correlation equal to 0.
- If there is no edge between cells i and j and the correlation is non-zero, then the nonzero correlation is due to all other data.

Formulation of the Bayesian graphical model

y: observed data

$y \mid \beta, \Sigma \approx N(\beta, \Sigma)$

Σ: Hyper Inverse Wishart Distribution

$e_{ij} \approx Bernoulli(p_{ij})$: edge between two nodes

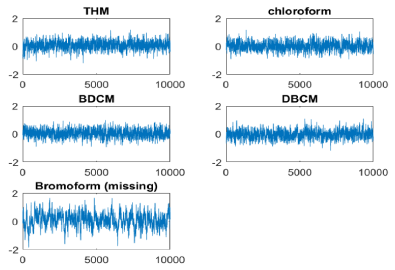
Prior probabilities on edges (representing existence of partial correlations)

- We give high prior probabilities to edges when two nodes have a close relationship.
- Such prior probabilities are still subjective, so they should be determined based on scientific knowledge to minimize subjectivity.

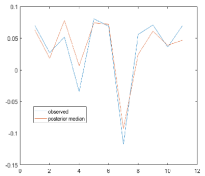
Results

95% Confidence Intervals of Human Data

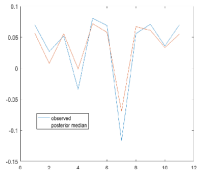
THM	Chloroform	BDCM	DBCM	Bromoform (missing)
0.0855	0.0234	0.0481	-0.0235	0.0878
(-.459,.652)	(-.536,.576)	(-.490,.578)	(-.571,.559)	(-.856,.998)



Validation of the Proposed method (Cross Validation)



Human data 1–THM, 2–Chloroform, 3–BDCM (assumed to be missing), 4 DBCM



Human data 1 THM, 2 Chloroform, 3 BDCM, 4–DBCM(assumed to be missing)

Comments on Results

- Estimate of missing value has more variation.
- When BDCM (or DBCM) is assumed to be missing, the posterior median of predicted values of human BDCM (or DBCM) is close to the observed value of BDCM (or DBCM).
- The patterns of the posterior medians are similar to those of the observed data.

Discussion and Future Directions

- We followed NRC (2014) recommendations on using Bayesian analysis and specifically investigated methodology proposed by DuMouchel and Harris (1983) and Jones et al. (2009).
- We proposed a new Bayesian method and validated recovery of missing human dose-response using Jones et al. (2009) data
- We will use simulation studies to validate our new method and consider its application to additional real data sets.
- We will consider extending the idea of graphical model to the area of combining DNA or RNA sequence data generated from different species.
- We will also consider application of the methodology to more data-poor examples that are more common in IRIS assessment work.

The views expressed in this poster are those of the authors and do not necessarily represent the views or the policies of the U.S. Environmental Protection Agency

U.S. Environmental Protection Agency
Office of Research and Development

Appendix E

Committee Findings Regarding 2014 Recommendations

Item	Chapter	Recommendations from 2014 NRC Report[a]	Finding	Evidence
1A	2	EPA needs to complete the changes in the IRIS process that are in response to the recommendations in the [2011] NRC formaldehyde report.	The 2014 report reviewed and encapsulated recommendations from the 2011 report, so the present committee focused its review on assessing progress made in implementing recommendations made by the 2014 report.	Workshop presentations, posters, and discussion Recent IRIS documents (such as plans, protocols, and assessments) and tools.
1B	2	[EPA needs to] specifically complete documents, such as the draft handbook, that provide detailed guidance for developing IRIS assessments. When those changes and the detailed guidance, such as the draft handbook, have been completed, there should be an independent and comprehensive review that evaluates how well EPA has implemented all the new guidance. The present committee is completing its report while those revisions are still in progress.	The revised handbook was not provided to the committee. EPA staff indicated that the handbook is under internal agency review and that its public release is expected in 2018. The agency further indicated that standard operating procedures might evolve as the IRIS program gains additional experience in performing systematic review and using emerging methods. The committee expects handbook revisions to be a continuing process, and EPA similarly characterizes the IRIS handbook as "evergreen." The committee observed that guidance for conducting newly planned IRIS assessments is contained in protocols, and EPA stated that some material currently in protocols might reside in the handbook. The amount of and need for overlap in the protocols and handbook could not be judged without seeing the handbook.	Slides 21–22 Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation)[b]
2	2	EPA should provide a quality-management plan that includes clear methods for continuing assessments of the quality of the process. The roles of the various internal entities involved in the process, such as the CASTs, should be described. The assessments should be used to improve the overall process and the performance of EPA staff and contractors.	IRIS management has taken multiple steps to ensure high-quality management, including the creation of expertise-specific work groups, systematic-review work groups, and other intermediate structures to improve the quality of the IRIS assessments. EPA has also used the SAB Chemical Assessment Advisory Committee to review IRIS assessments. Funding for contractors has decreased.	Slides 7–10, 151 The GAO audit of the IRIS program indicates that improvements in program management have occurred (Slide 10)
3	2	When extracting data for evidentiary tables, EPA should use at least two reviewers to assess each study independently for risk of bias. The reliability of the independent coding should be calculated; if there is good agreement, multiple reviewers might not be necessary.	EPA uses two people to extract data and, when needed, involves a third person to resolve conflicts. EPA also uses two people to complete the risk-of-bias evaluation.	Slide 39 Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation) (p. 17, line 1; p. 30, lines 18–20)[b]
4A	2	EPA should continue its efforts to develop clear and transparent processes that allow external stakeholder input early in the IRIS process.	EPA has adopted the process of soliciting public comment early through the release of assessment plans and protocols for public comment.	IRIS Web site Slides 24–25, 29
4B	2	[EPA] should develop communication and outreach tools that are tailored to meet the needs of the various stakeholder groups. For example, EPA might enhance its engagement with the scientific community through interactions at professional-society meetings, advertised workshops, and seminars. In contrast, greater use of social media might help to improve communications with environmental advocacy groups and the public.	Although this recommendation was not discussed specifically with EPA, the agency has worked in the past with the National Academies to identify experts that could provide input at IRIS workshops. The IRIS Web site provides features for sharing information via social-media tweets and Facebook. The calendar feature clearly indicates the schedule for public engagement events on IRIS assessments. EPA staff also discussed data- and tool-sharing with stakeholders to increase understanding and accessibility of systematic-review practices used to develop IRIS assessments.	IRIS Web site Slide 15

(Continued)

Continued

Item	Chapter	Recommendations from 2014 NRC Report[a]	Finding	Evidence
5	2	Similar to other EPA technical-assistance programs, EPA should consider ways to provide technical assistance to under-resourced stakeholders to help them to develop and provide input to the IRIS program.	This recommendation was not discussed specifically with EPA.	
6	2	The stopping rules should be explicit and transparent, should describe when and why the window for evidence inclusion should be expanded, and should be sufficiently flexible to accommodate truly pivotal studies. Such rules could be included in the preamble.	The issue of stopping rules was not specifically discussed, but the IRIS program has completed a rapid review of chloroprene, and this is consistent with this recommendation.	Systematic Review of Chloroprene Studies Published Since 2010 IRIS Assessment[c]
7	2	Regarding promotion of efficiencies, EPA should continue to expand its efforts to develop computer systems that facilitate storage and annotation of information relevant to the IRIS mission and to develop automated literature and screening procedures, sometimes referred to as text-mining.	EPA has made considerable progress in developing and upgrading the Health and Environmental Research Online (HERO) database and the Health Assessment Workspace Collaborative (HAWC) computer system to facilitate storage and annotation of data. Those systems are not subject to third party control. EPA is also using other software systems, including the Sciome Workbench for Interactive computer-Facilitated Text-mining (SWIFT) and related products for text-mining.	Workshop Demonstrations Slides 36, 92–116
8	2	More details need to be provided on the recognition and applications of expert judgment throughout the assessment-development process, especially in the later stages of the process. The points at which expert judgment is applied should be identified, those applying the judgment should be listed, and consideration should be given to harmonizing the use of expert judgment at various points in the process.	EPA has developed guided expert judgment to synthesize evidence on the basis of modified Bradford Hill criteria and for integrating evidence across data streams. The agency has developed working groups with expertise (such as PBPK) that can be applied to the assessment process. The draft chloroform protocol identified some situations when expert judgment will be used, including evaluation of studies to identify characteristics that indicate how informative the results are (p. 16, line 21) to perform outcome-specific study evaluations (p. 16, line 24).	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation)[b] Slides 8, 48, 69–86
9	3	EPA should establish a transparent process for initially identifying all putative adverse outcomes through a broad search of the literature. The agency should then develop a process that uses guided expert judgment to identify the specific adverse outcomes to be investigated, each of which would then be subjected to systematic review of human, animal, and in vitro or mechanistic data.	EPA has developed assessment plans that provide information about the scoping and problem formulation process. The plans are developed by using expert judgment and input from EPA regional offices and other stakeholders. Each assessment plan identifies the specific aims of the systematic review and the PECO statement.	IRIS Assessment Plan for Chloroform (Scoping and Problem Formulation Materials)[d]
10	3	For all literature searches, EPA should consult with an information specialist who is trained in conducting systematic reviews.	EPA staff indicated that they use an information specialist.	EPA protocol provides the name of the HERO librarian (see chloroform protocol, page vii);[b] that person has an MS in library and information science
11	3	EPA should include protocols for all systematic reviews conducted for a specific IRIS assessment as appendixes to the assessment.	The IRIS program has developed draft systematic-review protocols that are undergoing public comment before being made final. The protocols contain many of the elements identified by the 2014	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation), Appendix A[b]

			report as meeting best practices defined by IOM. Furthermore, the chloroprene reassessment included as appendixes the literature-search strategy and approaches for evaluating risk of bias in epidemiology and other human studies. The study objective, PECO statement, and methods used to search and screen the literature and evaluate studies were included in the main body of the report. That approach is consistent with this recommendation. The committee expects that some items found in the protocol can be addressed in the handbook. Including the analysis plan in the systematic-review protocols might lead to additional amendments to the protocol that could be minimized if they used a separate analysis plan.	Systematic Review of Chloroprene Studies Published Since 2010 IRIS Assessment[c]
12	4	The trajectory of change needs to be maintained.	The IRIS program has been responsive to the recommendations made in the 2014 report and is continuing the trajectory of change. The changes appear to have accelerated with the recruitment of new NCEA and IRIS leadership.	Workshop presentations, posters, and discussion Recent IRIS documents (such as plans, protocols, and assessments) and tools
13	4	The current process can be enhanced with more explicit documentation of methods. Protocols for IRIS assessments should include a section on evidence identification that is written in collaboration with information specialists trained in systematic reviews and that includes a search strategy for each systematic-review question being addressed in the assessment. Specifically, the protocols should provide a line-by-line description of the search strategy, the date of the search, and publication dates searched and, as noted in Chapter 3, explicitly state the inclusion and exclusion criteria for studies.	EPA systematic-review protocols contain descriptions of how evidence will be identified, including relevant search terms and databases to be queried. The protocols also include descriptions of inclusion and exclusion criteria.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation), Table 2 (p. 9), p. 12, Appendix A[b] Systematic Review of Chloroprene Studies Published Since 2010 IRIS Assessment[c]
14	4	Evidence identification should involve a predetermined search of key sources, follow a search strategy based on empirical research, and be reported in a standardized way that allows replication by others. The search strategies and sources should be modified as needed on the basis of new evidence on best practices. Contractors who perform the evidence identification for the systematic review should adhere to the same standards and provide evidence of experience and expertise in the field.	EPA systematic-review protocols contain descriptions of how evidence will be identified, including relevant search terms and databases to be queried.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation), Appendix A[b]
15	4	EPA should consider developing specific resources, such as registries, that could be used to identify and retrieve information about toxicology studies reported outside the literature accessible by electronic searching. In the medical field, clinical-trial registries and US legislation that has required studies to register in ClinicalTrials.gov have been an important step in ensuring that the total number of studies that are undertaken is known.	This recommendation goes beyond the scope of the IRIS program and therefore was not addressed by the committee during its review. Systematic-review protocols indicate that IRIS assessments include only publicly accessible, peer-reviewed information, which should be available through the databases identified by the IRIS program.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation) (p. 11, line 14)[b]

(Continued)

Continued

Item	Chapter	Recommendations from 2014 NRC Report[a]	Finding	Evidence
16	4	EPA is encouraged to use at least two reviewers who work independently to screen and select studies, pending an evaluation of validity and reliability that might indicate that multiple reviewers are not warranted. It is important that the reviewers use standardized procedures and forms.	EPA uses two persons to screen studies. Screeners use a structured form based on the PECO in DistillerSR.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation) (p. 12, lines 11–16)[b] Slide 39
17	4	EPA should engage information specialists trained in systematic reviews in the process of evidence identification, for example, by having an information specialist peer review the proposed evidence-identification strategy in the protocol for the systematic review.	The IRIS assessment team includes an information specialist. The specific tasks completed by that person are not clear. It is hoped that the handbook will clearly define the roles that the person has in the IRIS process.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation)[b]
18	4	EPA should encourage and support research on reporting biases and other methodologic topics relevant to the systematic-review process in toxicology.	EPA is supporting and encouraging research through its collaborative efforts described at the workshop. The committee expects EPA research in this field to emerge as the IRIS program continues to develop expertise in systematic-review method development.	Slides 79, 91, 149, 145, 150
19	4	EPA should continue to document and standardize its evidence-identification process by adopting (or adapting, where appropriate) the relevant IOM standards described in Table 4-1. It is anticipated that its efforts will further strengthen the overall consistency, reliability, and transparency of the evidence-identification process.	Appropriate tools and methods for evidence identification were described and are being used.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation)[b] Workshop presentations
20A	5	To advance the development of tools for assessing risk of bias in different types of studies (human, animal, and mechanistic) used in IRIS assessments, EPA should explicitly identify factors, in addition to those discussed in this chapter, that can lead to bias in animal studies—such as control for litter effects, dosing, and methods for exposure assessment—so that these factors are consistently evaluated for experimental studies.	The draft chloroform protocol describes the domains to be considered in the evaluation of epidemiology studies and animal toxicity studies. Domain ratings and their descriptions have also been provided. EPA also presented heat maps of risk-of-bias analyses for studies performed by the IRIS program.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation) (Tables 5–6)[b] Slides 53, 55
20B	5	Likewise, EPA should consider a tool for assessing risk of bias in in vitro studies.	The 2014 report noted that few tools were available for assessing risk of bias in in vitro studies. Fully developed tools that meet the needs of the IRIS program are not available. EPA is exploring adaptations of existing tools for its purpose.	Slide 78
21A	5	When considering any method for evaluating individual studies, EPA should select a method that is transparent, reproducible, and scientifically defensible.	EPA has adopted systematic-review methods that are transparent and scientifically defensible.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation)[b] Systematic Review of Chloroprene Studies Published Since 2010 IRIS Assessment[c] Slides 50–63

21B	5	Whenever possible, there should be empirical evidence that the methodologic characteristics that are being assessed in the IRIS protocol have systematic effects on the direction or magnitude of the outcome. The methodologic characteristics that are known to be associated with a risk of bias should be included in the assessment tool. Additional quality-assessment items relevant to a particular systematic-review question could also be included in the EPA assessment tool.	EPA is using and adapting risk-of-bias tools appropriately.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation)[b] Systematic Review of Chloroprene Studies Published Since 2010 IRIS Assessment[c] Slides 52–63 Posters D-4, D-5, D-9
22	5	EPA should carry out, support, or encourage research on the development and evaluation of empirically based instruments for assessing bias in human, animal, and mechanistic studies relevant to chemical-hazard identification. Specifically, there is a need to test existing animal-research assessment tools on other animal models of chemical exposures to ensure their relevance and generalizability to chemical-hazard identification. Furthermore, EPA might consider pooling data collected for IRIS assessment to determine whether, among various contexts, candidate risk-of-bias items are associated with overestimates or underestimates of effect.	EPA is supporting and encouraging research through its collaborative efforts described in the workshop.	Slides 145, 149
23	5	Although additional methodologic work might be needed to establish empirically supported criteria for animal or mechanistic studies, an IRIS assessment needs to include a transparent evaluation of the risk of bias of studies used by EPA as a primary source of data for the hazard assessment. EPA should specify the empirically based criteria it will use to assess risk of bias for each type of study design in each type of data stream.	EPA has adapted existing risk-of-bias tools for its use. Draft protocols describe the domains to be considered in the evaluation of epidemiology studies and animal toxicity studies. Domain ratings and their descriptions have also been provided. EPA also presented heat maps of risk-of-bias analyses for studies performed by the IRIS program. Tools have not been developed for mechanistic studies.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation) (Tables 5 and 6)[b] Slides 53, 78
24	5	To maintain transparency, EPA should publish its risk-of-bias assessments as part of its IRIS assessments. It could add tables that describe the assessment of each risk-of-bias criterion for each study and provide a summary of the extent of the risk of bias in the descriptions of each study in the evidence tables.	EPA presented example heat maps of risk-of-bias analyses for studies performed by the IRIS program. The heat maps have been included in a recent assessment.	Systematic Review of Chloroprene Studies Published Since 2010 IRIS Assessment (Figure 2)[c]
25	5	EPA should develop terminology for potential sources of bias with definitions that can be applied during systematic reviews.	EPA has adapted existing risk-of-bias tools for its use. The draft chloroform protocol describes the domains to be considered in the evaluation of epidemiology studies and animal toxicity studies. Reporting bias was not included as a domain for epidemiology studies, and its omission is not consistent with standard systematic-review methods for assessing risk of bias.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation) (Tables 5–6)[b] Slides 55, 57

(Continued)

Continued

Item	Chapter	Recommendations from 2014 NRC Report[a]	Finding	Evidence
26	5	Funding sources should be considered in the risk-of-bias assessment conducted for systematic reviews that are part of an IRIS assessments	EPA documents funding source, but it is unclear how the data are used.	Workshop discussion
27A	5	EPA should contact investigators to obtain missing information that is needed for the evaluation of risk of bias and other quality characteristics of included studies.	Investigators are contacted on a case-by-case basis that depends partly on the expected effect of the missing data. IRIS systematic-review protocols also indicate that decisions are made on an assessment-specific basis. If the information is not reported, it is generally not useful to reach out to the study authors. However, if missing study details could change confidence in study conclusions, efforts should be made to contact the study authors. Outreach to study authors is documented and considered unsuccessful if researchers do not respond to multiple e-mail or phone requests within a reasonable period.	Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation) (Table 6 (p. 25); p. 18, line 41)[b]
27B	5	The committee expects that, as happened in the clinical literature in which additional reporting standards for journals were implemented (Turner et al. 2012), the reporting of toxicologic research will eventually improve as risk-of-bias assessments are incorporated into the IRIS program. However, a coordinated approach by government agencies, researchers, publishers, and professional societies will be needed to improve the completeness and accuracy of reporting toxicology studies in the near future.	This recommendation goes beyond the scope of the IRIS program and therefore was not addressed during this review.	
28	5	The risk-of-bias assessment of individual studies should be carried forward and incorporated into the evaluation of evidence among data streams.	The results of the evaluation of individual studies are a critical component of the current evidence synthesis processes and integration frameworks. Risk of bias is one factor that EPA uses to determine an overall study confidence rating for epidemiology and animal toxicity studies. High- or medium-confidence studies are favored for quantitative dose–response analysis.	Slides 66, 54, 71–73, 81
29	6	EPA should continue to improve its evidence-integration process incrementally and enhance the transparency of its process. It should either maintain its current guided-expert-judgment process but make its application more transparent or adopt a structured (or GRADE-like) process for evaluating evidence and rating recommendations along the lines that NTP has taken. If EPA does move to a structured evidence-integration process, it should combine resources with NTP to leverage the intellectual resources and scientific experience in both organizations. The committee does not offer a preference but suggests that EPA consider which approach best fits its plans for the IRIS process.	The IRIS process continues to use a guided expert judgment process, but structured sets of categorical criteria for decision-making within that process are more explicitly defined.	Slides 67, 79–86

30	6	EPA should expand its ability to perform quantitative modeling of evidence integration; in particular, it should develop the capacity to do Bayesian modeling of chemical hazards. That technique could be helpful in modeling assumptions about the relevance of a variety of animal models to each other and to humans, in incorporating mechanistic knowledge to model the relevance of animal models to humans and the relevance of human data for similar but distinct chemicals, and in providing a general framework within which to update scientific knowledge rationally as new data become available. The committee emphasizes that the capacity for quantitative modeling should be developed in parallel with improvements in existing IRIS evidence-integration procedures and that IRIS assessments should not be delayed while this capacity is being developed.	EPA illustrated its use of meta-analysis of human and animal studies for evidence integration. Bayesian methods are being explored to help to characterize uncertainty and to combine evidence to identify hazard. New methods and assays are increasingly being evaluated quantitatively.	Slide 130 Posters provided examples that show how EPA uses new approach methods as part of a chemical assessment process
31	6	EPA should develop templates for structured narrative justifications of the evidence-integration process and conclusion. The premises and structure of the argument for or against a chemical's posing a hazard should be made as explicit as possible, should be connected explicitly to evidence tables produced in previous stages of the IRIS process, and should consider all lines of evidence (human, animal, and mechanistic) used to reach major conclusions.	The 2017 Toxicological Profile for Benzo[a]pyrene shows well-developed evidence tables that support the structured narrative and conclusion regarding carcinogenicity. For other effects, the evidence is described as ranging from "strongest evidence for human hazards" to "less robust evidence." Workshop discussion and the chloroform protocol show progress in template development. EPA staff stated that the approach to standardization of hazard descriptors for noncancer effects is being tested and discussed in the agency.	Slides 80–86 2017 IRIS Toxicological Profile for Benzo[a]pyrene[e] Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation)[b]
32	6	Guidelines for evidence integration for cancer and noncancer end points should be more uniform.	Although EPA has not developed these guidelines, the issue goes beyond the IRIS program with respect to agency procedures. However, the IRIS program has developed frameworks for evidence integration and is testing and discussing how conclusions should be summarized.	
33	7	EPA should develop criteria for determining when evidence is sufficient to derive toxicity values. One approach would be to restrict formal dose-response assessments to when a standard descriptor characterizes the level of confidence as medium or high (as in the case of noncancer end points) or as "carcinogenic to humans" or "likely to be carcinogenic to humans" for carcinogenic compounds. Another approach, if EPA adopts probabilistic hazard classification, is to conduct formal dose-response assessments only when the posterior probability that a human hazard exists exceeds a predetermined threshold, such as 50% (more likely than not likely that the hazard exists).	Progress has been made. Quantitative toxicity values are restricted to studies with strongest conclusions for a human health effect (for cancer, a descriptor of *Known*) or a moderately strong conclusion for a human health effect (for cancer, a descriptor of *Likely*). Criteria are not provided for inclusion of studies that are considered on a case-by-case basis when a weaker conclusion regarding a human health effect (for cancer, a descriptor of *Suggestive*) is reached. IRIS has not produced final descriptors for noncancer effects and mechanistic studies other than review and application of PK/PBPK models.	Systematic Review of Chloroprene Studies Published Since 2010 IRIS Assessment[c] 2017 IRIS Toxicological Profile for Benzo[a]pyrene[e] Slides 131–133

(Continued)

Continued

Item	Chapter	Recommendations from 2014 NRC Report[a]	Finding	Evidence
34	7	EPA should continue its shift toward the use of multiple studies rather than single studies for dose-response assessment but with increased attention to risk of bias, study quality and relevance in assessing human dose-response relationships. For that purpose, EPA will need to develop a clear set of criteria for judging the relative merits of individual mechanistic, animal, and epidemiologic studies for estimating human dose-response relationships.	Progress has been made toward using multiple studies or end points and comparing multiple candidate toxicity values. IRIS assessments provide one or more candidate toxicity values for use by risk managers. The IRIS program considers the quality of studies when deciding which studies will be advanced for quantitative dose–response modeling; studies rated as having medium or high confidence will be advanced for dose–response considerations. Other study attributes—such as relevance of a species to humans, relevance of an exposure route, and susceptibility—might also be considered. EPA is developing new tools for making and visualizing comparisons. EPA recognizes that there is no one-size-fits-all sets of criteria for inclusion of mechanistic studies, but the criteria for evaluating PK/PBPK models and how they are applied in dose–response and toxicity-value determinations are a good start.	Slides 62, 130–135, 142–146 2012 IRIS Toxicological Review of Tetrachloroethylene[d] Workshop demonstrations of HAWC and SWIFT
35	7	EPA should use formal methods for combining multiple studies and the derivation of IRIS toxicity values with an emphasis on a transparent and replicable process.	IRIS has begun to develop and apply tools in response to this recommendation. EPA presented two demonstrations for meta-regression and Bayesian approaches that showcase the agency efforts. EPA has not presented criteria for when and how new tools should be used. Tool development and application will be a continuing process that requires sustained resources and continued capacity-building.	Slide 140 Case studies provided for alternative dose estimates (posters D-2, D-10)
36	7	EPA should clearly present two dose-response estimates: a central estimate (such as a maximum likelihood estimate or a posterior mean) and a lower-bound estimate for a POD from which a toxicity value is derived. The lower bound becomes an upper bound for a cancer slope factor but remains a lower bound for a reference value.	EPA indicated that this approach is now standard procedure. Several examples were presented that show comparisons between BMDs and BMDLs and demonstrate how key studies compare with other supporting studies	Slides 134, 135; posters
37	7	As the IRIS program evolves, EPA should develop and expand its use of Bayesian or other formal quantitative methods in data integration for dose-response assessment and derivation of toxicity values.	Demos show the beginning stage of IRIS efforts on applications of Bayesian methods. EPA has not yet developed criteria for when and how new tools should be used. New research is under way to address New Approach Methods, such as data-mining, cheminformatics, high-throughput exposure modeling and toxicokinetics, and visualization tools.	Case studies (Poster D-10) Slides 136, 139, 140, 143–146

38	7	Uncertainty analysis should be conducted systematically and coherently in IRIS assessments. To that end, EPA should develop IRIS-specific guidelines to frame uncertainty analysis and uncertainty communication. Moreover, uncertainty analysis should become an integral component of the IRIS process.	Efforts are beginning with development of model-averaging approaches and adoption of WHO/IPCS guidance for reporting toxicity values and their uncertainty. IRIS specific guidance has yet to be developed because tools and approaches remain under development.	Cooper et al. (2016)[g] Slides 137, 138 Case studies (Poster D-9)

[a]NRC. 2014. Review of EPA's Integrated Risk Information System (IRIS) Process. Washington, DC: The National Academies Press.

[b]EPA. 2018. Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation) [CASRN 67-66-3]. EPA/635/R-17/486. Integrated Risk Information System, National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC.

[c]Orme-Zavaleta, J. 2018. Response to the Request for Correction (RFC). Letter to Robert Holden, Liskow & Lewis, New Orleans, LA, from Jennifer Orme-Zavaleta, Principal Deputy Assistant Administrator for Science, Office of Research and Development, Washington, DC, January 25, 2018; Attachment 1. EPA Response to the Denka Performance Elastomers (DPE) Request for Correction of the Toxicological Review of Chloroprene (CAS No. 126-99-8) In Support of Summary Information on the Integrated Risk Information System (IRIS); Attachment 2. Systematic Review of Chloroprene [CASRN 126-99-80] Studies Published Since 2010 IRIS Assessment to Support Consideration of the Denka Request for Correction (RFC). January 2018 [online]. Available: https://www.epa.gov/sites/production/files/2018-01/documents/epa_repsonse_to_mr._holdren_jan_25_2018_complete.pdf [accessed February 9, 2018].

[d]EPA. 2017a. IRIS Assessment Plan for Chloroform [CASRN 67-66-3]. EPA/635/R-17/330. Integrated Risk Information System, National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC.

[e]EPA. 2017b. Toxicological Review of Benzo[a]pyrene. EPA/635/R-17/003Fa. Integrated Risk Information System, National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC [online]. Available: https://cfpub.epa.gov/ncea/iris/iris_documents/documents/toxreviews/0136tr.pdf [accessed February 14, 2018].

[f]EPA. 2012. Toxicological Review of Tetrachloroethylene (Perchloroethylene) [CAS No. 127-18-4]. EPA/635/R-80/011F. U.S. Environmental Protection Agency, Washington, DC. February 2012 [online]. Available: https://cfpub.epa.gov/ncea/iris/iris_documents/documents/toxreviews/0106tr.pdf [accessed February 27, 2018].

[g]Cooper, G.S., R.M. Lunn, M. Ågerstrand, B.S. Glenn, A.D. Kraft, A.M. Luke, and J.M. Ratcliffe. 2016. Study sensitivity: Evaluating the ability to detect effects in systematic reviews of chemical exposures. Environ. Int. 92-93:605-610.